DOCUMENT RESUME

ED 141 387 TM 006 330

AUTHOR Fox, G. Thomas, Jr.
TITLE Limitations of a Standard Perspective on Program Evaluation: The Example of Ten Years of Teacher Corps Evaluations.
PUB DATE [Apr 76]
NOTE 102p.; Paper presented at the Annual Meeting of the American Educational Research Association (60th, San Francisco, California, April 19-23, 1976)

EDRS PRICE MF-$0.83 HC-$6.01 Plus Postage.
DESCRIPTORS *Data Analysis; Decision Making; *Evaluation Methods; *Federal Programs; Higher Education; *Models; Policy Formation; *Program Evaluation; *Teacher Education; Teachers Colleges
IDENTIFIERS Higher Education Act 1965; *Teacher Corps

ABSTRACT

Six evaluation studies on Teacher Corps are first described and then analyzed. The two evaluation perspectives used for analysis were: (1) the standard evaluation paradigm as described by Astin and Panos and (2) the evaluation paradigm suggested by Parlett and Hamilton as "illuminative evaluation". The standard evaluation paradigm describes three main components: student inputs, student outputs, and educational operations; and finds causal relationships between these components. In contrast to the standard evaluation paradigm, the Parlett and Hamilton paradigm includes: (1) a study of the entire program: its rationale and evolution, its operations, achievements and difficulties; (2) an examination of the program within the school or educational milieu; and (3) a combination of methodological strategies for describing and analyzing a program in order to help "illuminate" salient program features. Results of this study indicate that, if one agrees with Astin and Panos that "the fundamental purpose of evaluation is to produce information which can be used in educational decision-making", then these evaluation studies have failed. They failed not because they did not follow the standard evaluation techniques reviewed by Astin and Panos, but because they did follow these procedures and used the standard perspective of evaluation. The challenge for evaluations, then, seems not be apply these same evaluation procedures better, but rathr to use alternative perspectives on evaluation. (Author/MV)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED141387

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

LIMITATIONS OF A STANDARD PERSPECTIVE ON PROGRAM EVALUATION: THE EXAMPLE OF TEN YEARS OF TEACHER CORPS EVALUATIONS

G. Thomas Fox, Jr.

University of Wisconsin-Madison
225 N. Mills St.
Madison, WI 53706
608/263-4670

A version of this paper was presented at the 1976 AERA Convention, April, 1976 at San Francisco, California

TM006 330

Abstract

This paper is a critical analysis of past evaluation studies on Teacher Corps. Six evaluation studies are first described and then analyzed. The two evaluation perspectives used for analysis were: 1) the standard evaluation paradigm as described by Astin and Panos in Educational Measurement (Second Edition, 1971) and 2) the evaluation paradigm suggested by Parlett and Hamilton as "illuminative evaluation" (1976). The standard evaluation paradigm is well known. It includes describing three main components: 1) student inputs, 2) student outputs, 3) educational operations, and finding causal relationships between these components. In contrast to the standard evaluation paradigm, the Parlett and Hamilton paradigm includes a) a study of the entire program: its rationale and evolution, its operations, achievements and difficulties, b) an examination of the program within the school or educational milieu, and c) a combination of methodological strategies for describing and analyzing a program in order to help "illuminate" salient program features.

Most studies ranked quite high using criteria from the standard evaluation perspective. Conclusions and recommendations from these studies, however, when compared to actual agency policy decisions are shown to be very low in their applicability to changes in policies or in the understanding of process or replication of successes of Teacher Corps. Thus, an analysis of these studies forces one to a rather uncomfortable conclusion that the studies were highly competent in terms of the standard evaluation paradigm but of little value to three major purposes to program evaluation: policy decision making, replication or understanding.

On the other hand, the studies ranked very low when using the perspective of Parlett and Hamilton. For example, data collection procedures were limited, the perspectives were limited and there was no description of the actual processes being used for training.

Thus, the results of this study indicate that, if one agrees with Astin and Panos that "the fundamental purpose of evaluation is to produce information which can be used in educational decision-making", then these evaluation studies have failed. They failed not because they did not follow the standard evaluation techniques reviewed by Astin and Panos but because they did follow these procedures and used the standard perspectives of evaluation. The challenge for evaluations, then, seems not to apply these same evaluation procedures better but rather to use alternative perspectives on evaluation that will offer more appropriate information for making educational policy decisions. The challenge, therefore, is upon educational evaluators to identify more effective perspectives of evaluation that are alternatives to the one creatively reflected in these six studies.

Acknowledgment

Jim Steffensen, Chief of Program Development, Teacher Corps, must take some credit for the existence of this paper. Without Jim Steffensen's repeated requests for the paper and his concern that the review of these studies be shared with Teacher Corps personnel, this paper would never have been completed. I must thank him for his interest and concern while I curse the difficulty of the task.

Table of Contents

	Page
Introduction	1
Review of Studies	5
RMC I	6
Harvard	10
RMC II	15
Abt	19
Corwin	26
Marsh	40
Discussion	61
Conclusion	78
Recommendations	80
Footnotes	86
References	89

Introduction

(while you and i have lips and voices which
are for singing and to kiss with
who cares if some one-eyed son of a bitch
invents an instrument to measure spring with?

e e cummings[1]

Educational evaluators may understand this verse from cummings in two ways. First, it may be considered to reflect the feelings of many educators that program evaluation is unfair, unnecessary and a waste of time. Second, we may be understanding this verse as an indictment on certain practices that try to measure what cannot be measured or are too limited in their perspective for their own purposes.

The purpose of this paper is to address the second of these concerns with the focus being upon federal program evaluations. Since, in many cases, there is seldom the option to evaluate or not evaluate, one of the most difficult challenges to the profession is how to best evaluate an intentional program that, like spring, consists of change, life, growth, expectations and broken dreams. More specifically, the following question is asked: Are observed inadequacies of federal program evaluation due to not following standard evaluation procedures or are current standard evaluation procedures themselves inadequate for many evaluation purposes?

Although the question is not a new one, it has usually been discussed on either an ideological/theoretical level[2] or as responses to specific large scale studies such as the Coleman or Jencks reports.[3] The discussion here will use as an example the past ten years of evaluation

studies that have been conducted on one federally funded program. The program chosen for this discussion in Teacher Corps, a program initiated under the Higher Education Act of 1965 and currently being funded at 37.5 million dollars a year.[4]

For the purposes of this discussion, Teacher Corps should make a useful example for educational evaluators from a variety of persuasions and interests. There are three main reasons why Teacher Corps should be of considerable interest.

First, the complexity of Teacher Corps has made it a special challenge to many different evaluators. During the ten years of its existence, Teacher Corps has undergone nine evaluation studies from various perspectives. These perspectives have included studies funded through the Office of Management and Budgeting,[5] studies funded privately through the National Education Association and the Ford Foundation[6] and on studies that have been initiated and sponsored by the agency itself.[7] Since Teacher Corps has been a challenge to evaluate, a number of creative procedures were used by some highly professional agencies.[8] In other words, the particular studies that will be reviewed in this paper have originated from a variety of sources and represent some interesting attempts at federal program evaluation.

The second reason that makes Teacher Corps an interesting case for evaluators is that it has continually shown a lively ability to be flexible in re-directing and changing its policy, focus and direction. Not only has Teacher Corps survived in an environment where most programs have not (at one time it was one of about ten programs associated with the Bureau of Educational Personal Development, now it is the lone survivor), but, judging from recent Presidential and

Congressional advisory councils, it appears that Teacher Corps may grow in the near future.[9] Currently, in fact, it is undergoing a major shift in emphasis from pre-service to in-service training and from the rhetoric of a service agency to that of demonstration projects.[10] Since Teacher Corps has been lively, flexible and dynamic in its policy changes, it is possible to compare policy changes to evaluation findings. This fact makes it possible to reflect upon the extent to which the evaluation studies affected Teacher Corps policy decisions without having to assume, as is often the case, that a bureaucratic inertia has made it impervious to change. It is an agency whose policy has changed which should make it possible to examine the extent to which the evaluation studies have helped determine these changes.

In addition to the fact that Teacher Corps has been an interesting challenge to evaluators and that it has been relatively flexible in determining new policy and emphases, the third reason that may make Teacher Corps interesting to a variety of educational evaluators is that the field of teacher education itself is in such a need for more insight and understanding based upon empirical findings.[11] Teachers, teacher educators and the general public have long agreed that there is much more to be learned about how teachers develop into being better teachers. Thus, Teacher Corps is in the enviable position of being in a field where even a little insight and illumination could be considered a dramatic increase in our understanding.

Simply put, Teacher Corps should provide program evaluators an interesting case study into the strengths and weaknesses of current

evaluation procedures. Not only has Teacher Corps survived for over ten years and undergone a number of evaluations performed by highly accredited evaluation teams and agencies, it also should have been an extremely fertile ground for program evaluation. It has been a complex and challenging program to evaluate, it has shown and is continuing to show itself to be a flexible decision-making agency and it has operated within a field whose practitioners have long recognized is in need of insights and empirical evidence. Considering these points, past evaluations of Teacher Corps should have been valuable both to internal program policy decisions and to furthering professional understanding of the processes of teacher education.

This paper will argue that the evaluation methodologies used, even when applied by some of the more creative, insightful studies, was much too limiting either for illuminating policy decisions or for furthering our professional knowledge. The argument will be made by briefly reviewing the six evaluation studies on Teacher Corps, discussing their major findings, relating these findings to Teacher Corps policy decisions and to the profession's understandings of the processes of teacher education. The studies will then be discussed using two different perspectives of evaluation, one representing a standard perspective of evaluation,[12] one representing an alternative perspective of evaluation.[13] After this discussion, a conclusion will be reached about the appropriateness of standard evaluation paradigms for federal program evaluations. Implications will then be drawn for future evaluation studies with specific suggestions for changes in policy and procedure for evaluating federal programs.

Review of Studies

Six evaluation studies on Teacher Corps will be reviewed in this section. In chronological order, these six studies are:

1. An Assessment of Teacher Corps. Report UR, 2 volumes. Bethesda, Md.: RMC Incorporated (Robert Crosby, director). August, 1970. Grant no. OEC-0-099018-4647. (Will be referred to as RMC I.)

2. "The Teacher Corps: A Case Study in Evaluation," Chpater IV, in The Role of Education in the Bureau of Educational Professional Development. Boston: Center for Educational Policy Research, Harvard University. December, 1971. pp. 106-153. (Will be referred to as Harvard).

3. "Teacher Corps," Volume I and Part G, Volume II of Report UR-172. Full Scale Implementation of a Process Evaluation System for Programs of NCIES. Bethesda, Md.: RMC Incorporated (Robert Crosby, director). November, 1972. (Will be referred to as RMC II.)

4. Innovation and Change: A Study of Strategies in Selected Projects Supported in NCIES. 4 volumes. Cambridge, Mass.: ABT Associates. December, 1972. Grant no. OEC-0-71-3714. (Will be referred to as Abt.)

5. Corwin, Ronald G., Reform and Organizational Survival: The Teacher Corps as an Instrument of Educational Change. New York: Wiley-Interscience. 1973. (Will be referred to as Corwin.)

6. A Study of Teacher Training at Sixth-Cycle Teacher Corps Projects. Volume I. Methodology and Findings. July, 1974. A Study of the Effectiveness of Sixth-Cycle Teacher Corps Graduates. Phase II. October , 1974. Teacher Corps Graduates in Highly Innovative Roles in Education: A Follow-Up Study. Phase II. February, 1975. U.S. Office of Health, Education and Welfare, Office of Education, Office of Planning, Budgeting and Evaluation. Contract no. OEC-0-73-5174. (These three reports will be referred to as Marsh.)

The review of each of these studies will include: a) a brief introduction to each study including who sponsored the study, the purposes and major questions of the study, b) a description of each study

including the major variables identified, the data collection procedures used and the data analyses performed, c) the major findings of the study, d) the relation or impact of these findings to actual (or potential) Teacher Corps policy decisions, and e) the relation of these findings to professional understandings into processes of teacher education.

RMC I (1970)

Brief description of RMC I Study.[14] This study was performed under contract to the Office of Education and was prepared specifically for Teacher Corps. The purpose of the study was to determine the extent to which Teacher Corps was achieving its stated goals in the field, and to attempt to identify program characteristics that contribute to success of failure. The major questions of this study were a) What were the legislative goals and short term (process) objectives of Teacher Corps?, b) To what extent was Teacher Corps achieving these goals and objectives? and c) What were the program structural characteristics that contributed to success or failure in achieving these goals?

Study Design of RMC I. Since the major purpose of the RMC I study was to assess the degree to which Teacher Corps achieved its objectives, a discrepancy evaluation design was used. Seven general legislative goals were identified and these were further broken down into forty-five process objectives that reflected Teacher Corps guidelines to funded projects. Ten sample projects were identified and evaluated according to the extent to which the legislative goals and process objectives were reached.

Since a second purpose was to identify program characteristics that contribute to success or failure, six program "structural characteristics" were used to describe each program. Each of these program "structural characteristics" had two descriptive categories. This meant that there were six dichotomous terms that could be used to describe any particular program. The six "structural characteristics" or program variables were:

1. program management - defined as either principally governed by the school district or principally governed by the university
2. team structure - defined as either "loose" where teacher trainees had no direct classroom responsibility, or "tight" where teacher trainees had direct classroom responsibility
3. program length - either 14 months or 24 months
4. program type - either graduate school training or undergraduate training
5. size of university - either small or large identified through enrollments in the School of Education
6. size of area served - either urban or rural and included the school district enrollment

Although the ten sample programs were chosen in order to get as wide a representation of the six "structural characteristics" as possible, it was a limited sample that could represent only a fraction of the possible combinations of these six dichotomous variables.

Data collection involved a great amount of time and effort in identifying the forty-five process objectives through interviews and review sessions with Teacher Corps personnel. Assessments of project success in achieving each of these goals and process objectives were made through on-site visits, interviews of project participants

and questionnaires. For each project, a description was drafted and a rating (of 3 possible levels) was given by two observers/interviewers. Differences in ratings and descriptions were resolved in confrontations between the two observers. Percentages of process objectives reached were computed for each legislative goal by project and for all the projects. Conclusions about the effectiveness of Teacher Corps were based on the percent of overall achievement for each of the seven legislative goals. Recommendations for improving Teacher Corps performance were made using these percentages of project effectiveness and the corresponding differences in the six structural characteristics.

Findings of RMC I. Teacher Corps had made clear progress in recruiting and selecting highly motivated student teachers. Furthermore, the placing of these student teachers in schools serving low-income areas did provide for more small group instruction and for more non-instructional services to the community itself. There was little evidence, however, that the more ambitious legislative goals were being achieved. Individual Teacher Corps projects were having little impact on schools or universities. Local school practices were not being significantly changed and Teacher Corps training practices were being isolated and not adopted by the cooperating universities. Although closer relationships were being developed between school districts, universities and communities during the life of the projects, there was little indication that this coordination would extend beyond the life of Teacher Corps funding. Community service was dependent upon individual efforts of the teacher trainees themselves and, thus,

was disjointed and not significantly effective. Broader professional and public awareness of the accomplishments of local projects was not observed at all, nor was there any evidence that others were being encouraged to adopt the successful practices.

Relation of RMC I findings to Teacher Corps policy. Recommendations from the RMC I study included a change in Teacher Corps priorities. At the time of the RMC I study, emphasis of Teacher Corps was to 1) train teachers, 2) provide services to the schools and 3) change university training. The RMC I study suggested this be realigned to: 1) change university training, 2) change schools, 3) provide service to schools and 4) train teachers. There were no concrete suggestions on how to perform this change and, thus, there was little impact of these recommendations on Teacher Corps policy. Later, as will be seen, the original Teacher Corps emphasis, as described in RMC I, was altered considerably. The other major recommendation from RMC I was to adopt the list of forty-five process goals that had been developed and to use these goals in an extensive internal management system for monitoring local project successes and failures. Again, since there were no concrete suggestions on how this would be actually accomplished (other than by questionnaires), this, too, did not affect operational decisions of Teacher Corps. Five years later, however, a management system was incorporated into the Teacher Corps program.[15]

Relation of RMC I findings to professional understandings of teacher education. The arid nature of the six structural variables (e.g. graduate vs. undergraduate training) combined with the minimal

ratings and description of the achievement of process goals make the illumination from this study extremely limited. Other than findings that Teacher Corps projects were having difficulties achieving the more ambitious legislative goals (regardless of individual energy and performance at the local level) there was little that would be considered to be an insight into teacher education.

Harvard (1971)

Brief description of Harvard study. This case study of Teacher Corps was part of a larger study on the Bureau of Educational Professional Development (BEPD). It was performed by the Center for Educational Policy Research under contract with the National Advisory Council on BEPD. The purpose of this study was to assess the role of evaluation in the operations of the programs within BEPD. Teacher Corps was used as a sample federal agency where the focus was on how Teacher Corps used evaluation in making, remaking, and changing its policy. There were three basic questions that the Harvard study tried to answer: 1) Is evaluation of such programs possible? (and if so, how?), 2) Is Teacher Corps the sort of program anyone in government wants to learn from?, and 3) What does the experience of Teacher Corps suggest about the evaluation of discretionary, experimental and reform programs elsewhere in education?

In a sense, then, the Harvard study had a similar purpose as this report. It was aimed at describing Teacher Corps, stating what had been found about Teacher Corps and suggesting what may be useful for

future Teacher Corps policy making. The difference between the Harvard study and this report, however, is that the Harvard study interpreted "program evaluation" as basically a singular procedure "resembling a reasearch-development-field testing and evaluation cycle" and was seen as "one way of learning about the effects of educational reform and change." (underlining added here) This report will be considering evaluation in a wider context and, in fact, will be suggesting an alternative perspective to the one assumed in the Harvard study.

Study design of Harvard Study. This study was an historical, anecdotal report of how Teacher Corps was created by the legislature and how it operated as a decentralized federal agency. The purpose was to describe Teacher Corps official policy towards "evaluation" as well as the actual practice of Teacher Corps in using "evaluation." As stated before, the focus was on a certain perspective of evaluation and the question was to what extent evaluation as so defined was useful for Teacher Corps purposes. The irony is that the procedures used by the Harvard study to answer this question were not "evaluative" using their own definition.

The methods of data collection used in the Harvard Study were difficult to ascertain but, judging from the spirit and content of the report, data were mainly collected through interviews, memos, anecdotal accounts of dinner meetings and luncheon speeches, some informal reviews of legislative reports, Teacher Corps guidelines and an awareness of (at that time) ongoing Teacher Corps evaluation

studies. Interviews of central Teacher Corps staff (those based in the Washington, D.C. office) were used.

The procedures for analyzing this information also were unclear but, again interpreting from the spirit of the report itself, was probably very informal and idiosyncratic to the writers of the report. Little cross-checking or verification of the data seems to have taken place.

Reference to the data collection and analysis of this study must mention the liveliness of the writing style. The report is interesting, lucid and smoothly written, full of personal anecdotes, quotes from memos, legislative reports and proceedings and newspaper articles. The style invites serious mistakes of oversimplification and quite often seems to be written with a Freudian tongue in cheek ("Galbraith's ideas [in March, 1964], may have been seminal but the Teacher Corps probably was born at a luncheon meeting in Washington, some months later"[16]). Nevertheless, the style also invites the reader's attention. The historical account of the legislative and administrative beginnings of Teacher Corps are very informative.

Findings of Harvard Study. A general conclusion of this study was that Teacher Corps was essentially a political effort based upon long established assumptions about social change (e.g. reduce inequality through improved education). The political nature of Teacher Corps was indicated by continuous changes and revisions in its aims and methods. The study suggested that trainees were in contradictory roles of change agent, student and servant to the schools. It questioned the assumption

that university change produces change in schools and concluded that the key role for local projects was the school principal. The study emphasized the purse-string authority of Teacher Corps through proposal guidelines and grant making but concluded that the local control of projects made it questionable that evaluative information could be generated or utilized. For example, it was critical of Teacher Corps' attempts to apply technical support systems such as "cumbersome" time-flow charts and cost-benefit analyses.

Generally, the Harvard study concluded that program evaluation (as they had defined program evaluation) may not be possible for the Teacher Corps context. Although the rhetoric of Teacher Corps suggested it was an experimental program interested in program evaluation, in the view of this study, evaluation was of low priority and "one might almost say it is foreign to the character of the program itself."[17] This conclusion is found in the summary of the Harvard study. The summary begins by acknowledging that Teacher Corps (even by 1971) had existed longer than any other BEPD program. It then goes on to say:

> If any single point emerges from this study, it is the fact that the political and administrative climate in which Teacher Corps has been situated has not been particularly encouraging to evaluation. Before we conclude that Teacher Corps was incorrect in not supporting more evaluation efforts, or better ones, we must ask whether -- given the organizational environment -- it wasn't following a sensible course in paying no attention to evaluation.[18]

Relation of findings to Teacher Corps policy. As interesting as these findings are, they have had little to do with policy changes in

Teacher Corps. Other than the rather obvious findings about the contradictory roles of interns, the importance of school principals and the question about the efficacy of the focus upon university change, there was little in this study that would relate to specific Teacher Corps policy. The conclusion that Teacher Corps was not interested in evaluation efforts is clearly false, judged both by its support of other studies reviewed in this paper as well as its current concern with documentation and demonstration.

Relation of findings to professional understanding of teacher education. One may expect that the limited view of program evaluation and the rather flippant spirit of the report (e.g. that reform programs cannot be "evaluated") would restrict its usefulness to the profession. If, for example, the Office of Education held a view of evaluation less restrictive than that proposed by the Harvard study (i.e. if they held a notion of evaluation that depended less upon a priori objectives and predetermined treatments), then the major conclusion of the Harvard study would not be applicable.

Nevertheless, regardless of the flippant tone, or inapplicability to program policy or faulty conclusions on Teacher Corps intentions, this report should be necessary reading for teacher educators. Teacher education is a political as well as an educational enterprise. The political negotiation that accompanied the original drafting of the Teacher Corps legislation and that has continued to shape its existence and definitions of problems is clearly a feature of all teacher education programs. As limited as this report is to specific program

policy, the first section of the report provides a great deal of illumination into the political nature of teacher education and provides some insight into the extent to which shared notions of social change are consistent with designs for educational reform.

RMC II (1972)

Brief description of RMC II. The RMC II study was a two year study prepared for the National Center for the Improvement of Educational Systems (NCIES) performed by the Resource Management Corporation. At the time, Teacher Corps was one of twelve NCIES programs. The major purpose of the RMC II study was to develop and then use a process evaluation system for all twelve programs. Although the aims of the RMC II study were to reach conclusions on funding/refunding decisions for all twelve NCIES programs and to monitor all NCIES ongoing project and program activities, Teacher Corps did get special referral because it was different in many respects from the other eleven programs.

Study design of RMC II. There were two major variables used to compare the twelve NCIES programs: participant characteristics and project characteristics. Participant characteristics included the percentage of those recruited from outside the field of education, the percentage of participants expecting degrees, the percentage of those receiving state certification, the percentage of those admitted to college on a provisional basis or for whom entrance requirements were waived. The project characteristics included the percentage of projects that changed state certification requirements, the percentage of projects budgeting for project continuation after withdrawal of federal funds and

the percentage of projects serving schools eligible for Title I ESEA assistance. In addition to those variables identified for all twelve NCIES programs, there were also unique sets of participant and project characteristics for each of the twelve programs.

Data were collected by two surveys: one addressed the common characteristics of all NCIES programs and one was program specific. Both surveys were mailed to local program directors who were responsible for distributing the surveys. Survey data was cross-tabulated and descriptive data was reported in tables. A discrepancy analysis was performed using specific program goals and successful performance was determined by two types of summaries: one summary stated percent of measures successfully established by each project (this was used to rate each project), the other stated the percent of projects that met each measure successfully (this was used to rate each NCIES program).

Findings of RMC II. In comparison with other NCIES programs, Teacher Corps did very well. For example, most projects from other programs did not recruit from outside the field of education, Teacher Corps did; most (75%) Teacher Corps interns received state certification where the average of the other NCIES programs was only 29%; most (88%) Teacher Corps projects served schools eligible for Title I ESEA assistance where the overage of the other programs was only 48%.

The report of the project characteristics specific to Teacher Corps included recruitment, objectives, advisory councils and evaluation. Recruitment involved many different role groups and resulted in trainees being selected from a variety of racial/ethnic backgrounds. Objectives

were focused upon changing university training procedures rather than improving or changing the school system. Although advisory councils existed, they seldom met. Almost all Teacher Corps projects said evaluation was in progress, that measurable objectives had been set and that evaluation was being managed by the project staff (notice the contradiction between these findings and the predictions of the Harvard study). Formal evaluation activities (e.g. evaluation design, data collection and data analysis) were being conducted only by outside consultant and college staff. Informal evaluation activities (e.g. comments, opinions and perceptions) involved other project participants. Other interesting survey results were that most (72%) participants felt that the local projects had caused them to do things differently. Thirty percent felt they had new attitudes towards teaching, 33% felt they had an increased regard for children, most felt that practical teaching experience was the strongest feature of their project and most felt that the weakest project feature was the lack of communication between participants.

It was reported that, for the discrepancy analysis, Teacher Corps had developed "the most extensive, detailed and comprehensive set of program conditions of any of the twelve NCIES programs."[19] RMC II also reported high minimum levels for satisfactory performance where over two-thirds of the projects equaled or exceeded 80% of the goals stated by the Teacher Corps program. The study concluded that these findings "indicate that projects are closely adhering to the majority of the program conditions, thus meeting the intent of the guidelines

and the Congressional legislation."[20] Recommendations suggested for Teacher Corps included: a) increasing the flexibility of the interns' training program, b) placing specific emphasis on community participation and responsibilities of advisory council, c) taking action to ensure continuation of project developments, d) continuing the stress on practicum but placing more emphasis on the community component and e) not letting interns substitute for classroom teachers or being responsible for regular classrooms.

Relation of findings to Teacher Corps policy. Since Teacher Corps was the only NCIES program to have survived, one could conclude that these findings were extremely important. On another level, however, the recommendations and findings have not been helpful for determining policy. The recommendations, for example, had no reference to context or suggestions for helpful procedures on how to perform the actions suggested. The figures for goal achievement gave no indication of why some projects were achieving these goals and some projects were not. Thus, although the comparative results were healthy for Teacher Corps' existence and the discrepancy analysis suggested successful implementation of guidelines, the complete lack of evidence of the procedures being used to achieve high performance left Teacher Corps with little to base future program policy.

Relation of findings to professional understanding of teacher education. The generality of the recommendations for NCIES programs and the rigid adherence to measuring only the perceived achievement of programmatic goals supplies no illuminative information to the profession at large.

ABT (1972)

Brief description of the Abt study. This study was prepared for the Office of Planning, Budgeting and Evaluation, Office of Education by Abt Associates, Inc. and, like the RMC II study, was directed towards all NCIES programs (of which at that time Teacher Corps was one of eight). The focus of the Abt study was on the processes of institutional change, that is, how innovations are (or are not) implemented by institutions and incorporated into educational practices. The intent of the Abt study was to supply policy relevant information useful to planning future "change-oriented" programs. The expressed purpose of the study was "to provide programs a framework for helping to determine the most favorable mix of project organization, innovation strategy and training content for reaching the desired goals."[21] The purpose was not to measure the impact of each project but to identify patterns within projects which facilitated or constrained change.

Some of the major questions of this broadly based study were a) What are the constraints a project director faces and the strategies used to overcome these constraints?, b) What are the influences that affect the effectiveness of programmatic aims directed through different institutions? and c) What effects may be produced by variables that can be manipulated by OE either through program guidelines or through grant award procedures?

Study design of Abt study. As may be expected from the wide-ranging purposes of this study, the study design was complex and involved. Initially, a conceptual scheme for describing project variation was developed. Projects were then examined in order to

verify and enrich that conceptual scheme. Using the scheme, strategies were identified that seemed most promising and productive. There were five major variations or dimensions (each representing a continuum) upon which a project was placed. These five variations included: 1) the service orientation of the project (ranging from logical extensions of school structure to providing new personnel expected to shift school priorities), 2) the training focus (from a university-based, non-traditional portal school environment to workshop programs organized outside the framework of the university setting, 3) the practicum experience (from nontraditional roles for trainees to specific practice of particular skills), 4) the centrality of the projects (from integrally built in to the normal functioning of the university to separated from and involving little contribution from the university), and 5) the project's organization structure (from being diffuse in decision making to involving trainees and staff together in planning goals and strategies).

Data was collected through case studies. Thirteen sites were chosen as being exemplary by their respective programs (two of these were Teacher Corps sites). Two interviewers (a "senior" and a "junior" interviewer) were trained. The training included a seminar in institutional change, a preliminary one day structured visit conducted and then reported to colleagues. After the structured visit, a draft of a plan for a week-long visit was drawn and approved by the study director. Each plan included tentative hypotheses to be investigated, a full schedule of interviews for the first three days and a written confirmation of the interview schedule from the site.

All week-long visits were completed within a six-week period. Although the report designs of each visit were idiosyncratic to that visit, there were five content areas that had to be covered in all the case studies. These were: 1) project descriptions, including verbal description of IHE, LEA, community and individuals, 2) identification of strategies used to effect institutional change, including both "conscious" and "unconscious" strategies, 3) the organizational dynamics for change including who made decisions, an organizational chart of the project and the interviewers' judgment of how change could actually have occurred within the participating institutions, 4) the impact of the project as perceived by project participants including the interviewers' own subjective interpretations of subtle areas of potential impact and 5) other unique features that may be idiosyncratic to the project site and may also affect the replicability of the strategy. Several drafts, reviews and editings were made of each case study report with each site project director being involved in periodic reviews and editings of the reports.

Unfortunately, as extensive, open and detailed as these case studies were, they were reported but not actually used in the discussion of the results, nor in the recommendations. (This mistrust of the qualitative case study data may be attributed to their perception that case studies had never before been used in large scale evaluation studies.) Rating scales were used for cross-site analysis with the findings, conclusions and recommendations of the study being based upon the analyses of these ratings. Items for the rating scales were based upon literature

on change theory and were written by the interviewers after their case study visits. Thirty-seven items were developed and each project was rated on the items using a Likert 7 point scale. (Items later were cross-indexed with the case study reports.) The thirty-seven variables were clustered by rater judgment, with the clusters being justified using a factor analytic procedure. Ten construct variables were then identified and construct means for each project were computed by transforming negative scores, adding all scores and dividing by the number of scales used in the construct variables. The analytic procedures included determination of means, standard deviations and pairwise correlations between all 37 scales and the ten construct variables as well as partial correlations, multiple regression and causal analyses.

One interesting feature of the study that was discussed but, unfortunately not pursued in any detail, was a mapping sentence. The mapping sentence was designed to specify possible ranges of context and effects. It was a non-quantitative procedure using the judgment and perceptions of those constructing the sentence. The general mapping sentence followed this format: projects of [_____], given contexts of [_____] cause institutional changes of [_____]. Within each bracket were a number of possible descriptions. The "full mapping sentence" was more complicated than this general form because of further expansion of the descriptive categories within the brackets. The project characteristics were considered to be independent variables, capable of manipulation by OE or project management (e.g. amount of funding, duration

of funding, pre-service or in-service). The context variables were certain characteristics of the participating institutions that could not be directly manipulated (e.g. attitudes towards change, support of a specific educational innovation). The change variables were the major concerns of both OE and project personnel. These included the permanence, magnitude and impact of the changes as well as where the changes occurred.

Findings of Abt study. The case studies on the two Teacher Corps projects were rich both in detail and in professional wisdom. For example, one of the case studies (a thirty-three page report) commended the project on its programmatic accomplishments but questioned its impact. The accomplishments were documented as new techniques for teaching reading and writing, new types of teacher recruits, modular instruction, team teaching, differentiated staffing and introducing English as a second language. However, the report went on to call these changes "cosmetic" since there was no fundamental alteration in the problems associated with racism, educational bureaucracy or teacher-centered classroom instruction. The report goes on to suggest that this may be an endemic problem to the programmatic approach to educational change. A final section suggested that the project did nothing to change the tendency towards de facto segregation but, instead, seemed to "buy time" by diffusing dangerous situations. The project did not deal with racial discrimination, community dissatisfaction with the status quo, maldistribution of resources or ineffective educational bureaucracies.

> Given the problem addressed, this Teacher Corps project has brought about the intended changes in the style and even the content of schooling in the classroom it serves. Yet, the consequences of such innovations do not appear to be systemic. Although it may take four or five years to adequately measure the results of Teacher Corps community impact, new staffing patterns, curricula and concepts of instruction do not seem to deal with the primary problems faced by young children of low-income families.[22]

Compared to the tone and spirit of the case study report, the findings produced by the scales, construct variables and analytic procedures are relatively tame and uninteresting. The comparison between the two Teacher Corps projects and the other eleven projects showed many similarities in the 37 scaled variables except that a) Teacher Corps personnel had more training, b) Teacher Corps projects had less institutionalization into the university and c) Teacher Corps projects had more risk and controversy than most other projects. Probably the most important result of these comparisons, however, (and one not pursued in the study) were the large differences between the two Teacher Corps projects. The differences between the two Teacher Corps projects on all but one of the ten constructed variables was greater or at least as great as between either Teacher Corps project and any project from a different program.

Although all the results and findings of the Abt report that depended upon the construct variables and data analytic procedures must be highly suspect, there were three findings that were also collaborated in the case studies. These findings were: 1) regardless of the value placed on change by the project staff or the presence of change training in the curriculum, it was the organizational structure

of the project that was most closely related to implementing an innovation, 2) the importance of "boundary personnel" within the cooperating organizations could not be overemphasized for change to occur and 3) the universities were extremely resistant to change.

Relation of findings to Teacher Corps policy. As rich as this report is in the description of two Teacher Corps projects and as conscientious in its attempt to specifically address policy decisions, there is little here that could be useful for Teacher Corps policy. The wisdom expressed in the case studies, for example, provides no clues about what could be done other than the ameliorative programmatic efforts already undertaken and accomplished.

Although the problems and paradoxes of programmatic funding were identified (e.g. technical innovation versus long-range institutional change), suggestions for alternative actions were scant. Simply put, this study was a good example that understanding the dilemmas inherent in pursuing certain goals was not the same thing as providing helpful suggestions.

The suggestions made, based upon the constructed variables, were both inappropriate and at least questionable as to their validity. On the other hand, the creativity with which the Abt study was pursued, especially the care with which the case studies were developed and, possibly, the sentence mapping technique for identifying manipulative programmatic variables may prove helpful to Teacher Corps in fulfilling its recent mandate for documentation, dissemination and demonstration.

Relation of findings to professional understanding of teacher education. First, the case studies of the Abt report should become standard reading to those interested in changing teacher education. They do illuminate the problems and provide concrete evidence of the dilemmas that confront a project funded for a finite period of time to produce specific innovations. Second, the data gathering and interpretive procedures mentioned above (the case study process and the sentence mapping technique), used with more confidence, may be helpful to those interested in documentation and evaluation of specific teacher education programs. Third, the review of literature on institutional change is extensive and should be informative to many organizationally illiterate teacher educators. Fourth, both the perception of and evidence for the importance of institutional and political arrangements when encouraging innovations in teacher education may be illuminating to the teacher education profession. In short, although there are few specific recommendations that are useful to Teacher Corps policy, there is much in the Abt study that need be communicated to the profession at large.

Corwin (1973)

Brief description of Corwin. The Corwin report was a two-year study funded through National Education Association and the Ford Foundation. The study was headed by Ronald G. Corwin, a sociologist at Ohio State University and is the only Teacher Corps evaluation that has been published to date. The major purpose of this study was to gain insights into the possibilities for inducing institutional change.

The study considered Teacher Corps as an interesting case of a federally funded program designed to produce institutional change. It used its findings to a) generate some general propositions for subsequent research, b) test other specific hypotheses from sociological theory and c) speculate about general issues of federal intervention, organizational change and educational improvement. Unlike many of the other studies, the Corwin study was not necessarily designed to suggest specific changes in Teacher Corps policy.

The major questions of the Corwin study were: 1) What is the role of the federal government in reforming this country's educational system?, 2) To what extent can complex organizations be deliberately changed?, and 3) How can the professions be changed to improve service to low-income clientele?

The focus of the Corwin report is suggested by a quote from Sarason early in the study.

> In my experience, in practice, most explicit and implicit conceptions of change derive from the language and vocabulary of an individual psychology that is in no way adequate to changing social settings. The fact that one can be the most knowledgeable and imaginative psychoanalytic, learning or existentialist theoretician gives one no formal basis for conceptualizing the problem of change in social settings. The problem is simply not one to which these individual theories address themselves.[23]

Study design of Corwin. The assumption behind the Corwin study was simply that the capacities of organizations to change must be enhanced before improvements in education could occur. Thus, the variables chosen were a) variables related to organizational change and b) individual's attitudes towards change and initiative. Thirteen

variables were identified to indicate organizational characteristics or individual attitudes. Nine of these characteristics described individual attitudes (e.g. commitment to teaching, initiative and compliance in education, role conceptions, alienation, liberation, satisfaction, preferred teaching styles). One described the extent of innovation by estimating the extent of change at each project due to Teacher Corps involvement. The final two variables described the organizational setting through a college quality index and a regional modernization index.

Data collection procedures included extensive use of questionnaires, a one-week on-site visit to ten projects by a team of interviewers (in 1968-69), and return visits to five projects a year later (spring, 1970). Selected for return visits were the more highly rated projects. Each site visit included interviews of personnel at three cooperating schools as well as one comparison school. Observation at the sites included field notes and field diaries as well as extensive interviews.

For most of the thirteen variables, there were used multiple measurement scales and indices. These included a variety of Likert scale items and multiple choice items in addition to two especially creative indices for measuring a) iniative and compliance in education and b) professional and employee role conception. Initiative and compliance in education was assessed using nine hypothetical incidents based on conflict between teachers and administration. Respondents (trainees) were asked to anticipate their behavior and estimate the seriousness of the incident. The two scales, 1) what would be done and, 2) probable seriousness of the act in terms of official sanctions,

were combined to form a typology of initiative to compliance. Professional and employee role conception included four items to measure professional perspective. These four items were : a) regard for students, b) orientation to profession and colleagues, c) belief that competence is based on knowledge and d) belief that teachers should have decision making authority. The employee scale included loyalty to administration and emphasis on rules and procedures.

The major dependent variable, inducements in organizational innovation, were measured according to assessments of Teacher Corps involvement in affecting changes in a) school-college cooperation, b) classroom teaching techniques, c) school involvement in community and d) nature of college training. Items were assessed using open-ended interviews and content analysis of planned and unplanned changes. A five point weighting scale was used on five dimensions (methods, materials used, clientele, personnel involved, and creativity of activity) for each planned change. Unplanned changes were judged by an additional dimension of improved attitudes. The ratings were based on the assumption that complex changes requiring new or altered role relationships are more important than simple additive changes affecting a few individuals. Each of the forty-two (42) schools had a scale score computed from a product of the number of planned and unplanned changes times the weights assigned to each applicable dimension.

The procedure also included a review of the literature on effectiveness of teacher training, teaching roles and effective teacher characteristics and arrived at an "ideal type" of teacher for poverty children.

Comparisons were later made between this ideal type of teacher and the descriptive data on the trainees and practicing teachers.

A variety of data analyses were used to consolidate, compare and interpret the collected data. These included ANOVA tests to compare variability within as well as between Teacher Corps projects (e.g. on teaching effectiveness), correlations used to determine associations among dependent variables (to reduce number of independent variables), and regression analyses to explain the amount of variance due to newly constructed variables. Two case studies were also reported in order to show the different institutional contexts in which these projects were placed as well as to add more continuity to the discrete observations described above.

Findings of Corwin. The findings of this study reflected the broad spectrum of aims and the variety of information collected. Generally, the findings of the Corwin study related a picture of the Teacher Corps trainees (interns), discussed the efficacy of the change strategies being used by local projects (with special focus upon the trainee as change agent strategy), discussed the program outcomes in terms of university or school changes, and discussed the theoretical implications for organizational change with special focus on possible techniques for changing educational organizations through federal intervention.

The Corwin study, like the previous studies, found that Teacher Corps did attract different types of persons into teaching (for example more initiative, less loyal to administration, more liberal). The trainees defined their roles and objectives differently from other

teachers (stressing motivation, creativity and inquiry) and were more confident that children from low-income areas could learn. Although these trainees were not necessarily committed to a lifetime of teaching (in fact, Corwin later reports that they seemed both less confident and less appreciative of teaching classes on a day-to-day basis), they did approximate the proposed ideal type teacher more closely than did the experienced teachers. Furthermore, it was found that if trainees possessed one characteristic of the ideal type they tended to have all the others (i.e. the nine characteristics for the "ideal teacher" seemed to cluster for the trainees).

With these results, Corwin concludes that "probably the prime achievement of this program [Teacher Corps] was to provide a channel to attract this type of teacher."[24]

Evaluation of teaching competence was admittedly indirect and inferential, being constructed from personal ratings of cooperative teachers and supervisors. Four dimensions of good teaching were identified: a) innovativeness, b) understanding of pupils, c) rapport with pupils and, d) relationship with community. Results of these ratings showed more variability of the ratings within projects and between roles within these projects, rather than between projects.

In reporting the findings on the teacher trainees, Corwin also reported some of his own primary beliefs on the nature of schooling. These included his belief that it is almost impossible to separate teachers' classroom behavior from more general attitudes and values and the related contention that a person's general style and personal values would make a greater difference in the outcomes of schooling

than techniques. The on-site visits made to classrooms convinced him that "most classrooms are the grim, joyless places described by Silberman."[25] However, Corwin also expressed doubts that the humanistic, ideal type teacher would be appropriate or effective for low-income children (that is, for improving national test achievement). He then goes on to suggest that the need might be for a greater blend of the academic and humanistic styles with a greater range of teacher models. Finally, in discussing the trainees, Corwin attributes most of the programmatic changes he observed to their optimism, impatient criticism and sustained pressure rather than to any other single program feature.

This brings us to the major findings of the study which related organizational constraints and strategies for change to the implementation of educational innovations. Since this is the major area of Corwin's expertise and, indeed, was the major focus of the study, it is these results which are the most interesting, trustworthy and informative. Corwin reports, for example, that contrary to a belief in the efficacy of professional collaboration, increasing teachers' active participation in the project did not increase the likelihood of change. Indeed, participation by teachers was negatively correlated to a) the introduction of new techniques, b) cooperation between college and schools and c) the number of community projects undertaken. Likewise, it was suggested that the professors in schools of education were unprepared to deal with the poor. They were too academic and remote from school problems.

In Corwin's view, Teacher Corps was applying two general strategies for inducing institutional change. One strategy was to rely upon

technical authority and political power with little support of actual technical knowledge. The other strategy was to rely upon the power and influence of the inexperienced trainees as "change agents." Both of these institutional change strategies, Corwin concludes, were ineffective.

As reported in the Corwin study, the "trainee as change agent" strategy had its greatest impact upon the trainees. He reports, for example, that trainees became less disposed to exercising initiative, more disillusioned about school bureaucracy and administration, and less committed to teaching. The process of assimilation was described as being complex where some trainees (because of unwillingness to compromise or work within the system) removed themselves from a position of influence, other trainees (those who experienced least change) became "radicalized" out of the teaching profession during their two-year training, while a majority of all trainees became more convinced that basic changes in society were needed.

Teachers, on the other hand, showed a uniform decline in student orientation and an increase in loyalty to administrative rules, procedures and an emphasis on pupil control. For example, technological innovations were negatively correlated to the teachers' attitude that society needed basic changes.

Innovative, substantive change (as measured by participants' opinions) was judged to be minimal at either the university or school level. At the universities, for example, the most that was anticipated were a few new courses, some modification of course content,

some new teaching techniques to be employed such as role playing, games and microteaching, an increase in interest and attitude towards "disadvantage students" and more dialogue between professors and students. There was expected to be little improvement in college-school cooperation (for example, no classes were to be placed on site) and no changes in the university or departmental organization were expected. Similarly, any changes anticipated in the school due to Teacher Corps involvement were short-range changes at best. Trainees assisted teachers by improving the student-teacher ratio, taking up some of the work load and providing individual attention to some students. There were few anticipated incorporations of new ideas, teaching methods, or materials at the school sites and, again, no expected changes in organizational structure and few changes in school-community relations. In short, Corwin concludes that the trainees only supplemented the normal activities of the schools and were used only as an added source of manpower by the teachers.

An assessment of the consequences of introducing liberal newcomers into the teaching profession produced the following interpretations.

First, the major dimension of the environment contributing to innovation was the amount of cooperation between school and university. The combination of cooperating features that contributed to program goals was, however, far more complex than anticipated or imagined. For example, it was found that centralized decision making was more consistent with technological change than decentralized decision making. The control over funds, however, in these centralized organizations was

also a leverage for co-opting the program. In addition, although centralized school organizations did not necessarily resist technological change, they did react defensively to critical, liberal, change-oriented newcomers. Finally, as noted previously, the belief in the efficacy of power equalization was further challenged by the negative effect that teacher participation had in inducing innovations.

Corwin cautiously summarized these results into some general propositions about how to effectively introduce innovations into school organizations. These propositions focused upon a combination of variables that could be expected to induce change more consistently and efficiently than the independent change agent approach. The fundamental consideration, Corwin suggested, was the way local organizations were insulated from or integrated into, a larger context. Within this discussion, Corwin concludes that future research should give more attention to the conditions that enable schools to maintain a functional autonomy within an increasingly interdependent, changing society.

When the dependent variable was the amount of school-college cooperation, regression analysis showed two general results. First, the critical ingredient was the quality of the "boundary personnel" or those who were assigned laisson roles between school and university. Second, cooperation between school and university declined with teacher participation in the proposal itself.

The final chapter (Chapter Eleven) in the Corwin study discusses these results within a theoretical perspective. The view is sociological, the topic is techniques for changing and improving educational

organizations through federal interventions, the example is Teacher Corps. As a single piece of writing, this chapter is by far the most informative, creative and provocative forty pages of the thousands being reviewed in this essay. Chapter Eleven in the Corwin study should be required reading for all educators interested in implementing educational innovations.

Corwin interprets the main contribution of Teacher Corps as providing incentive and opportunity for more collaboration to occur among a variety of groups and organizations concerned with education. The critical innovation, from his viewpoint, is not curricular or instructional but is the fact of mutual participation by schools, universities and communities. This view further places the historical importance and significance of Teacher Corps as a deliberate attempt to establish a nationwide collaborative framework among organizations to attack an educational problem (e.g. improving the education of children from low-income areas). Corwin uses Teacher Corps as an example for inferring an interpretation of the reform process itself.

Five categories of variables are suggested for interpreting the reform process. These categories are:

1. the task structure of the organizations involved (e.g. goals, role structures and incentives),
2. the status system and identities of members (e.g. background, authority, competence, values),
3. economic resources available and how allocated,
4. distribution of power and internal processes used to resolve conflict, and
5. occupational setting.

Corwin suggests that of these five variables the distribution of power is the most critical since it is often the distribution of power that determines the priorities of the other four categories. Innovation, he contends, will be received and implemented according to the results of the power struggles and related political processes of bargaining and compromise.

Two models of organizations are then discussed. One is a rational model of organization, that is, that organizations are goal-directed entities where goals dominate the planning, the stating of alternative courses of action and the decisions. The other is a political model, that is, that organizational action is determined by external constraints, by previous commitments made in bargaining for resources and where official goals are only some of the possible sources of commitment and are themselves subject to compromise. Schools and universities, he suggests, can only be interpreted as political organizations. Negotiation, compromise, conflict and selective enforcement of conflicting norms are more closely associated with schools or universities than are the features describing rational organizations. The main point Corwin makes is that in a political organization, innovation must be expected to become ensnared in various internal and external power struggles that can only be understood in specific local circumstances.

With this in mind, the major variables that need be controlled by Teacher Corps, as an outside federal agency interested in inducing innovations, were reinterpreted. As interpreted by Corwin, these variables became a) the sources of incentive and constraint (where goals become the bargaining units), b) an identification of the critical targets for

change (for example, collaboration between schools and universities) and c) a centralized guidance role for Teacher Corps where evaluations and systematic monitoring help determine funding through a flexible discretionary budget. A lack of discretionary funds and the incapacity to gather information about local projects was suggested by Corwin to be major reasons that universities were so able to resist change.

Corwin concludes that Teacher Corps cannot be evaluated as if it were an agency that had a fixed, identifiable, achievable goal. Instead, faced with constraints beyond their control and little concentrated power to implement a program as planned or, even, permit sufficient internal control, Teacher Corps should be assessed upon the more relevant and appropriate consensus-seeking functions. These consensus-seeking functions would be a) sustaining the momentum for social reform, b) advancing alternative conceptions of innovation and c) facilitating debate and compromise.

Relation of findings to Teacher Corps policy. There are two ironies about the value of the Corwin report to helping determine Teacher Corps policy. The first irony is that, as provocative as the views and discussions of this study are, the few specific recommendations to Teacher Corps are relatively inane. The recommendations included a) that at least part of the intern training focus upon how to make changes in complex organizations and b) to select low status universities, select a critical mass of trainees and place them in positions of control. Nevertheless, one can, using the evidence of the Corwin study, easily go beyond the recommendations and offer the following suggestions. One, search for ways to include teachers in

the training process (there is no way that teachers can be circumvented when inducing innovation, regardless of the difficulties noted in the Corwin study). Two, require specific institutional and role arrangements that can be expected to support a greater openness to and propensity for innovation. Three, emphasize the documentation, dissemination and demonstration of the methods used to encourage both wider participation and greater consensus on educational issues. These are but a few of the specific recommendations that can be inferred from the Corwin study which suggests the second irony of this study. Although the Corwin study was the one study reviewed in this paper that had no intention of supplying Teacher Corps with specific recommendations for policy making, it has the greatest potential for helpful advice.

Relation of findings to professional understanding of teacher education. Like the Abt report, the Corwin study should be required reading for teacher educators, both to acquaint them with a perspective that is necessary for understanding the process of inducing innovations in the schools and as a provocative example of the political and institutional constraints that play upon efforts to implement educational reform. Unlike the Abt report, it is not the case studies that are so illuminating but, rather, the discussion suggesting a theoretical perspective of institutional change (Chapter Eleven).

It may be necessary to note that most of the findings of Chapter Eleven are not directly related to the statistical results of the study but are more related to the professional wisdom and interpretory powers of the author. As will be emphasized later in this essay, this inclusion of a personal and professional construct of meanings to the results

can be interpreted more as a praiseworthy asset rather than a weakness of the study. Again, it may be appropriate to note that this is the one study that is published and is, thus, easily accessible to the profession. In addition, there is a volume of responses to the Corwin study by four teacher educators and a sociologist that is also available.[26] If nothing else, this latter volume suggests the extent to which we teacher educators need be educated in the sociological perspective.

Marsh (1974)

Brief description of Marsh study. The Marsh study was performed under contract to the Office of Planning, Budgeting and Evaluation, Office of Education. Although it was begun under one corporation and finished under another, it was always under the control of its two principal investigators, David D. Marsh and Margaret F. Lyons. Possibly more than most other large-scale evaluation studies, it was also guided by the project monitor, Dr. Robert Hall of the Office of Planning, Budgeting and Evaluation. The aims of the Marsh study were set very high. It was to be a study of intern recruitment, training and performance that could or would, once and for all, determine whether Teacher Corps training made any difference in the trainees and in the students they taught. The study was undertaken in the fall of 1973, and followed one cycle of trainees for the two-year training and their first year of teaching. Up to now, it is the last effort to evaluate Teacher Corps performance by either the Office of Planning, Budgeting and Evaluation or any outside team of evaluators.[27]

There were two major goals of the Marsh study. These were to:

1) Identify and analyze those combinations of intern background characteristics and Teacher Corps program characteristics that are related to desired teaching skills and attitudes of interns at the end of their training, and

2) Assess the effectiveness of Teacher Corps graduates working with low-income/minority group children.

The first goal was the focus of Phase I ending in the volume titled A Study of Teacher Training at Sixth-Cycle Teacher Corps Projects. The second goal was the focus of Phase II, ending in the volume titled A Study of the Effectiveness of Sixth-Cycle Teacher Corps Graduates. In addition to the originally designed two phase study, there was a follow-up component to Phase II that attempted a) to re-analyze some of the data from Phase I and b) to describe educational activities that were involving Teacher Corps graduates other than public school teaching. (Teacher Corps Graduates in Highly Innovative Roles in Education: A Follow-Up Study, is the volume describing other educational roles of Teacher Corps graduates.)

Before introducing the methodologies and findings of the Marsh study, it is wise to point out the considerable expense of the Marsh study, the lofty aims of its intentions, the great amounts of energy that went into its construction and also its place as benefactor from the previous attempts to evaluate Teacher Corps. Both its successes and failures must be understood in the context of a large-scale effort to judge the effectiveness of Teacher Corps intern training. One characteristic of this effort was the variety of data gathering procedures used and the multivariate data analyses applied.

Study design of Marsh study. Although the stated intentions were for Phase I to function in an hypotheses-generating mode and Phase II to function in further testing and refining of these hypotheses, the two phases

actually had little in common. Phase I was designed to look for the most effective ways to train interns. Generally, this was done by searching for relationships between and among a) trainee background characteristics, b) training program characteristics and c) trainee classroom performance characteristics (called "exit characteristics"). Three of the goals of Phase II, on the other hand, were designed to assess the effectiveness of Teacher Corps graduates in their first year of training. Generally, this assessment was approached by contrasting the performance of Teacher Corps graduates (and their students) with that of control group (non-Teacher Corps) graduates. Although a fourth goal of Phase II was to relate background, past training, teacher performance and pupil performance, the only relationship between this goal and Phase I was to include the data describing the program characteristics for various canoninical and multivariate analyses with later performance data. There was no revision, clarification or further testing of the hypothetical statements made in Phase I from correlational analyses. Simply put, the Marsh study was really two studies (with the third volume reviewed here as an unrelated account of interns' experiences after graduation).

In addition, any underlying rationale or theoretical position of either phase had to be inferred. Phase I, for example, chose to define the teacher training process in terms of three major variables: a) intern background characteristics, b) program characteristics and c) performance or "exit" characteristics. Although the input, output, process approach to defining the problem may be obvious when given these three variables, it was never stated, nor were results from studies of similar perspectives

cited. There was, for example, no attention given the differences often noted in such studies between "manipulative" variables (those that can be expected to be changed through policy) and "non-manipulative" variables (those that cannot be expected to be changed). In simple terms, both Phases of the Marsh study were designed to produce tour de force statements about the effectiveness and impact of Teacher Corps training upon its recipients.

This superhuman aim was reflected in some almost superhuman efforts at data gathering through observations, interviews and questionnaires distributed through repeated on-site visits.

In Phase I, data were collected from all 20 Teacher Corps projects preparing elementary school teachers. Performance or exit characteristics were collected on sixty percent of the interns in a stratified random sample (by age and sex). Site visits of from one to two weeks were conducted at all sites both in the fall and spring semesters. Training for these visits required nearly three weeks. There were eleven questionnaires and about seven interview schedules used to gather data on the program characteristics above. Four instruments were used to "measure" performance or exit characteristics. These included a direct classroom observation guide, a guide for reconstructing trainee activities (both a questionnaire and interview schedule), supervisors' (team leaders) assessments of specific abilities and activities of each trainee, and a trainee self-rating form. The background data was collected through a thirty-item questionnaire.

The amount of time, effort and concern given the data gathering procedures for Phase I was, obviously, considerable as was the amount of data collected.

Dáta analyses for Phase I consisted mainly of factor analyses to reduce the number of variables. New program "factors" and performance (or "exit") factors were constructed using orthogonal and oblique rotations. Canoninical correlations were performed on the background, program and performance variables to determine "important trends" between program variables, background variables and performance. Multiple linear regression analyses were used to determine relationships between background and program factors with specific performance or "exit" characteristics. Relationships were drawn for the sample population of trainees as well as for three ethnic groups of trainees (Black, Chicano and White). Two-way analyses of variance and t-test comparisons were used to determine differences between graduate and undergraduate training projects on performance skills. Simple correlations were used to search for the extent to which Teacher Corps projects had had an impact on their host universities implementing a competency-based teacher education program.

For Phase II, a sample of Teacher Corps graduates (interns) was taken from the original pool of trainees and compared with a sample of recent graduates from non-Teacher Corps training programs. Both samples were teaching in public elementary schools (grades 2-6) with major responsibility for a specific group of children. Individuals from each sample were matched on grade level, pupil socioeconomic background, school district and years of teaching experience.

Performance variables upon which the two samples were to be compared were modified from the performance factors of Phase I. Twenty-three performance factors were reduced to seven teacher performance

clusters. These seven performance clusters were: 1) utilization of community resources, 2) action as a change agent in the school, 3) lesson planning and materials development, 4) classroom performance:affective tone, 5) classroom performance:teacher response strategies, 6) individualization of instruction and 7) perceptions of problems related to schooling. Data on teacher performance were gathered by a teacher questionnaire developed for the Marsh study and a previously developed classroom observation guide.

As stated before, the aims of Phase II were high and included not only comparison between Teacher Corps and non-Teacher Corps graduates on classroom behavior but also on their pupils' performance. The variables upon which pupils were measured included a) reading gain, b) self-concept gain, c) school attendance, and d) pattern of pupil classroom interaction. Data on these pupil variables were collected using the Metropolitan Achievement Test for reading ability, the Piers-Harris Self Concept Scale, school attendance records and the previously mentioned classroom observation guide for pupil interaction.

In addition to teacher performance variables and pupil variables, Phase II also identified four school climate variables that were considered as possible influences on teacher performance and effectiveness. These variables included a) the percent of low-income pupils, b) feelings of the teacher about other staff members and c) the school setting. This information was collected by the previously mentioned teacher questionnaire.

All data for Phase II of the Marsh study were collected in the fall and spring of the 1973-74 school year. The reading achievement and self-concept tests were given both in the early fall and late spring of the

school year. Classroom observations were conducted and teacher questionnaires distributed in the spring. There were twelve five-minute observations taken of each classroom over a consecutive two-day period.

As in Phase I, Phase II used a variety of data analytic procedures to "answer" specific questions. For example, one-way Analyses of Variance were used to compare the two samples of teachers (the Teacher Corps graduates and control group) on both the teacher performance measures and the pupil measures. A one-way analysis of covariance was used to compare the two samples on pre and post scores from the reading and self-concept tests. Canoninical correlations were used to identify teacher performance variables with pupil variables. Discriminant analyses were used to identify those teacher performance variables which best predicted "above-average" or "below-average" classroom mean scores. Simple correlation and discriminant analyses were used to search for relationships between teacher sex, teacher ethnicity and Teacher Corps program characteristics both with teacher performance variables and with pupil performance variables. Simple and canoninical correlations were used to determine relationships between school climate variables, teacher performance variables and pupil performance variables. A one-way analysis of variance was used to compare Teacher Corps and control group samples on each school climate variable. A Chi-Square analysis was used to compare the Phase II sample of Teacher Corps graduates with all interns from the Phase I study study on sex, ethnicity and project affiliation.

The follow-up study of Phase II was principally a descriptive account of Teacher Corps graduates' activities (and intentions) in the year following graduation. These descriptions were achieved in two ways. One was through a questionnaire sent to all Teacher Corps graduates (669) prepared as elementary school teachers. Approximately 50% (338) of these questionnaires were eventually returned. The questionnaire focused upon rudimentary job descriptions (e.g. extent to which schools served low-income pupils), trainees' reflections upon their Teacher Corps training experience (e.g. quality and usefulness of training to current position), trainees' attempts to locate jobs and their plans for the future. Results from the questionnaire were given in terms of the percent of those giving each type of response (e.g. 79% reported they worked in schools that served low-income children, 42% reported that their Teacher Corps training was "extremely useful").

In addition to the mailed questionnaire, there was an hour-long phone interview with approximtely 20 graduates who had "innovative roles in education." These roles included a) teachers of pupils with learning disabilities, b) teachers of English as a second language, c) reading, math and science specialists, d) special non-teaching jobs in education and e) jobs in institutions of higher education. Three to five graduates from each of these five job categories were selected on the basis of their job description in the questionnaire. The interviews were conducted in November and December, 1974. Although the expressed intention of the interviews was that "each job description would be somewhat independent of the unique experience of any one

graduate,"[28] the method used to come to these general job descriptions was not described. Descriptions of each job included a) general statements about task responsibilities and b) how those in that job category related their Teacher Corps experience to their current responsibilities. The number of graduates interviewed for each job category (three to five) and the purpose for these descriptions ("to identify patterns in graduates' professional growth"[29]) left these job descriptions with a dilemma. How to generalize from three to five cases when not reported in detail. That the resolution of this dilemma was not successful is clearly in evidence from the many times that phrases similar to "all teachers reported"[30] were stated when "all" referred to three, four or at the most five individuals. One time, for example, the statement was "several teachers found their Teacher Corps coursework helpful, several did not"[31] when there were only four persons who were interviewed. This attempt to make the job descriptions general rather than context specific affected the potential value in reporting and analyzing the phoned interviews.

These three separate phases of the Marsh study, taken together, represented a serious, professional attempt to produce useful information to Teacher Corps and to the teacher training profession. Not only was there an extensive use of surveys, questionnaires, interviews, observations and on-site visits but also all these techniques (with the possible exception of the case studies) were approached with a high regard for instrument validity and observer reliability. Furthermore, the data analytic procedures applied were some of the more sophisticated techniques available for correlational data. Although this reviewer has elsewhere criticized the categories used for identifying many of the

variables in the Marsh study, and the lack of a theoretical base,[32] nevertheless, any consideration of the Marsh study must include great respect for the amount of effort and diligence that went into the gathering and analyzing of data. Compared to the other studies reviewed in this paper, there is no doubt that the Marsh study was more ambitious in its intentions, more energetic in its approach and more conscientious in fulfilling its evaluation responsibilities as established by most current definitions.[33]

Findings of Marsh. Considering the zeal with which data were collected and the technical expertise through which the data were teased and analyzed, the irony of the Marsh study is the lack of illuminative findings. Admittedly, most of the evidence supplied through the Marsh study was correlative, that is, "strength of relationships" between certain variables and constructs of variables. Nevertheless, realizing that correlational data are never designed to give causal relationships but are, rather, designed to suggest whether certain relationships do or do not occur within certain contexts, the results are still relatively tame and uninteresting. As empirical as the Marsh study clearly was intended to be, most of the findings of Phase I can be interpreted as definitional problems.

For example, one of the "most important correlations"[34] between program factors and performance (or exit) factors was the strong inverse relationship (-.60) between the use of teacher competencies in the training program and the extent to which trainees felt that they were self-directed. The use of teacher competencies in a training

meant a) that certain skills were pre-determined as being necessary to successful completion of the training program, b) that these skills were determined before the training began, c) that all training activities were directed to achieving these skills, d) the mastery of these skills was assessed periodically by predetermined means and e) that as soon as a trainee passed the assessment, he or she could move on to the next skill. Considering that this describes activities underlying the phrase "the extent that competencies were used in training," the interesting result would have to be that the inverse relationship was "only" .60! Why would trainees feel self-directed in such a system? The answer to that question would have been illuminating. Partial answers would include: a) the trainees were duped into feeling self-directed for some reason, b) the use of competencies was not as efficient in practice as it was defined and intended by the trainers or c) there were other activities planned and performed in the interns' training that countered the determinism implied by the use of stated competencies. Here empirical evidence could have been informative, rather than merely a continuation of a logical relationship.

A similar example was the "finding" that the follow-up of interns' academic instruction (i.e. the extent to which university course work was tied into later classroom activities) was related to teacher participation in the project and to cooperative decision making between teachers and university professors. This finding, again, seems merely tautological where it would be far more illuminating to find out how follow-up was done without teacher participation or cooperative decision making (the relationships were .57 and .44 respectively)!

Empirical evidence would have been interesting only to the extent to which it did not corroborate the logical relationships.

Not all findings of Phase I of the Marsh study were of this variety. Some of the correlations tell what is already common knowledge, for example that positive attitudes held by university personnel towards a training program are inversely related to the extent to which teachers participate in the project (the old two-culture or practitioner vs. theoretician problem) or that the size of the school district is inversely related to the trainee's opportunity to bring about change in the school.

Nevertheless, there are a few results that warrant attention because they are difficult to explain, to understand or seem to contradict present professional wisdom. For example, the trainee's perception of the extent to which they have changed schools and the importance of these changes is inversely related to the amount of self-initiated contact with parents and to children-initiated interaction within the classroom. This suggests that the perception of important change brought about by a trainee is held by the trainee who is more autonomous and dogmatic in the classroom. Those who may be most innovative (e.g. initiate more contact with parents and allow children to initiate classroom interaction) give less importance to school change. The irony of this finding (i.e. those who make most changes see futility of these changes, those who make smaller changes see the changes as greater than they are) and the potential conflict this may raise within and between individuals should not be forgotten or easily disregarded by teacher educators.

Perhaps the clearest challenge to our professional beliefs comes from the relationships (more specifically the inverse relationships) found between trainees' performance characteristics. Before noting these relationships, it is appropriate to be reminded that these performance characteristics are not only the most clearly defined variables identified in the Marsh study, they also have been "measured" by some of the most reliable means used in this study. The characteristics were both recorded by trained observers and/or were self-reported as actions (rather than attitudes) by participants. There were twenty-nine performance characteristics (e.g. degree that instructional choices are given to pupils, intern initiates contact with parents: telephone calls) and the relationships between many of these actions are surprising and beg for futher interpretation. Some of the more surprising correlations are the following: a) trainee use of a broad range of resources for preparing lessons was inversely correlated to trainee's contact with parents and also inversely correlated to the trainee's acceptance and use of student ideas, whereas, b) the same characteristic (use of broad range of resources) was positively correlated to the use of an informal authority structure in the classroom and the extent to which trainees provided different learning activities for different children, furthermore, c) the trainee use of an informal authority structure was inversely related to attentiveness to children and also inversely related to an ability to communicate with pupils. In other words, the more trainees accepted and used student ideas, the less resources were used for instruction. The more informal the authority structure, the less

trainees accepted student ideas but a greater variety of resources were used. In short, in situations that teacher trainees usually identify as especially humanistic and capable of flexibility (informal rather than formal authority), they seem to behave more inflexibly and interact less with the students. In the Marsh report, there were many such findings about trainee performance that seemed to defy current assumptions or supposedly logical relationships between behaviors.

Another example of these provocative findings were the strong inverse relationships between the trainees following lesson plans and other activities that one would have expected to be related to lesson planning. Effective pupil diagnosis, the extent that trainees provided different learning activities for different children, asking open-ended questions, attending to children's responses and praising children all were negatively correlated with following lesson plans. In other words instead of lesson plans opening up the possibility for more appropriate responses to students, the more these plans seemed to cut off the trainee from the students.

There were many such interesting results of the correlative relationships between certain trainee performance activities. Although these results are capable of being interpreted from different perspectives (and, in fact, beg to be interpreted), there was no attempt in either the Phase I or Phase II reports of the Marsh study to deal with some of the more surprising or seemingly contradictory results. The one thing that was done, however, was to suggest two different "prototypes" of trainees. Type A introduced culturally relevant materials, did not

follow lesson plans, did not allow children to explore the room or select a work group. Type B focused on academic topics, allowed students to move about and select work groups, used lesson plans and used corrective feedback. The incongruity of the characteristics within each type was not addressed.

The remainder of the findings of Phase I tried to determine relationships between trainee background, program characteristics and intern performance characteristics. Generally this proved to be unsuccessful. Predicting trainee performance characteristics from trainee background and the program characteristics, for example, suggested only that one could predict a trainee's perception of the causes of poverty and poor reading skills. (These in turn were related to low-income and minority setting of the program.) Likewise, relating program factors with performance factors also provided little information. Out of sixty-five potential program factors, nineteen were correlated with at least one performance variable.

Generally, then, other than pointing out that little could be predicted about trainee performance from program or background characteristics, Phase I failed in its original intent to "identify and analyze those combinations of intern background characteristics and Teacher Corps program characteristics that are related to desired teaching skills and attitudes of interns at the end of their training."[35]

Phase II, however, one which set its goals even higher (that is, to assess the effectiveness of Teacher Corps graduates in working with low income/minority group children) was actually more successful in terms of that goal. Some interesting differences (and similarities)

were found between graduates trained by Teacher Corps and graduates trained elsewhere (the control group).

No differences were found in classroom performance characteristics. There were, however, two out-of-class performance characteristics that did differentiate Teacher Corps graduates from other graduates. Teacher Corps graduates a) developed more ethnically relevant curricula and b) used more community resources in teaching and initiated more contact with parents. There also was a difference in attitudes towards the causes of poverty and the causes of reading problems with Teacher Corps graduates attributing these causes to society and environment, other graduates attributing the causes to the individual.

There were no differences noted in pupil reading achievement. There were significant differences, however, in pupil self-concept scores. The pupils of Teacher Corps graduates had higher scores of self-concept about their physical appearance and higher scores of happiness and satisfaction with themselves. In summary, no differences in classroom interaction and teaching styles were observed betwen Teacher Corps and non-Teacher Corps graduates, no differences in pupil achievement were observed but there were differences in out-of-classroom activities (curricular design and use of community resources) and differences in pupil gain on certain self-concept subscales.

Teacher Corps graduates were also found to perceive that their views about curriculum and discipline were more divergent from other teachers than control graduates. Likewise, Teacher Corps graduates were less likely to receive ideas and curriculum materials from other staff members. Generally, it was found that for both Teacher Corps and

non-Teacher Corps graduates, those who "individualized" instruction (i.e. allowed children to work in small groups and/or individually and praised the children a good deal) were also likely to be those who a) did not obtain ideas or materials from other teachers and b) perceived their views as being not similar to other teachers. Considering the difficulties inherent in managing individualized instruction, this combination of attitudes might be considered to be self-defeating for inexperienced teachers in self-contained classrooms.

The attempt to relate training program features with graduate performance and pupil performance was not as successful. The few relationships found between program features and graduate performance features were the positive correlations between a) low team leader-trainee ratio and the use of community resources and b) strong team leadership during training and involvement in a change agent role in the schools. The few relationships between program features and pupil gain scores were even less informative (e.g. higher gains in pupil self-concept were related to projects with courses taught by white instructors).

Finally, correlational analyses between the teacher performance characteristics and pupil gain scores suggested that large pupil gains in both reading and self-concept were most closely related with a) teacher perception that reading development was primarily a function of pupil characteristics rather than teacher characteristics, b) a low level of teacher/pupil interaction, c) teachers acting to bring about change in the school, d) teachers initiating contact with parents and e) planning lessons with a team of teachers.

Findings in the follow-up portion of the Marsh study suggested that the majority (about 88%) had sought teaching jobs, had jobs in education (85%), mostly as teachers (77%) and were in schools serving low-income children (80%). Most planned to continue as classroom teachers in the following year (77%) and all who had jobs in education planned to remain in education. Most planned to stay in education after five years (82%) but, in general, saw themselves moving from classroom teaching (from 77% to 31%) to other educator roles (from 8% to 51%). Most rated highly the quality of their Teacher Corps training as well as the usefulness of that training to their current work.

Although the case studies of those in innovative educational positions were not formally summarized, the following were some generalizations made in the Marsh study. The more common themes stated about Teacher Corps training experiences were: a) the importance of the trainees' relationships with their training team leader, b) the importance of seeing a variety of instructional techniques, c) the value of gaining extensive experience in public schools, not necessarily in teaching but in understanding the operational aspects of schooling, d) their sense of personal growth during training, especially in dealing with frustration and ambiguity as well as taking personal responsibility for initiating action, e) the sense of commitment to educational innovation developed within a project related to a national effort, f) the problems within their training project, for example, little support and much disorganization, actually forced them to take initiative and organize, g) the value of their community involvement,

h) the view that only courses directly applicable to their special area of responsibility were useful, and generally, i) that their training experience was mostly portrayed as a "consciousness-raising experience, broadening their notion of education, schools, and teaching."[36]

A final theme of those in educational roles other than teaching was the energy toll of their work. Many were reported as feeling weary, isolated and overworked. Many felt a need to share their accomplishments with others but found little opportunity to do so.

Relationship of Marsh study to Teacher Corps policy. There are two important points to make about the relationship of the Marsh study to Teacher Corps policy. First, is the rather unfortunate circumstance that the completion of the Marsh study, directed at assessing the effectiveness of trainee preservice training, coincided with a change of Teacher Corps from primarily a training program for certifying new teachers to a training program for experienced teachers. Teacher Corps, after the Marsh study was completed, was quite different from the Teacher Corps evaluated by the Marsh study. Thus, one would expect that the Marsh study would not be particularly valuable as an aid in determining future Teacher Corps policy except as it may relate to the continued preservice component of Teacher Corps projects. (Four instead of thirty-five trainees are trained for certification while an entire school staff is the focus of the re-training component.)

The second point, however, is that the results of Phase II of the Marsh study (that there were some differences that favored both

the performance of Teacher Corps graduates and their impact on students) have been effective in supporting Teacher Corps at a time when decisions were being made about its survival. The directness of its findings and the simplicity of the Phase II design made the findings of the Marsh study valuable to Teacher Corps' current status in the Office of Education. To those evaluating Teacher Corps, the Marsh findings suggested that a) Teacher Corps was doing what it was supposed to (recruit and train minority members to teach in low-income schools), b) that its graduates were entering the educational field, primarily as teachers and c) that this training was having some impact upon the pupils in these schools. These results suggested to policy makers that, even though major changes in Teacher Corps emphases were being made (e.g. re-training experienced teachers, a single school focus and demonstration), that it may be able to carry off these responsibilities as well.

Nevertheless, when considering these two points about the Marsh study, one must still conclude that, as the study now stands, there is relatively little of value to Teacher Corps policy makers. Although the amount of data collected in Phase I was great, there is little interpretation of this data that can be used to identify possible agency alternatives in training nor even to judge whether certain training procedures are effective or not. Potentially, the most useful data (and certainly the most reliable) are the data describing trainee performance characteristics. Unfortunately, because of the grand design to determine relationships between program features and impact, this data, especially their ambiguity and incongruency, became relatively unimportant to the study. The contradictory results of the Marsh study,

pursued further, could suggest some interesting and illuminating consequences of certain training procedures used within the Teacher Corps context. Incongruencies such as a) contacting parents but b) not developing culturally relevant curricular materials, or c) following lesson plans but d) not providing different learning activities for different children have not been anticipated nor documented before. To resolve these incongruencies would be to address specific training nrocedures. For example, it could be suggested 1) since doing each activity well takes so much time, each training program should allow different trainees to focus on different features, or 2) that there are needed other training strategies that better address and combine these observed incongruencies. It must be noted, however, that the Marsh study makes no such mention of these possibilities. In fact, one could anticinate that, because of the lack of interpretation and attention given to such findings, most of the Marsh study will never be related to future Teacher Corps policy. It is now simply too difficult and tedious a task to be undertaken by the policy makers themselves.

Relation of Marsh study to professional understanding of teacher education. If the findings of the Marsh study for Teacher Corps policy can be considered to be potentially but not realistically useful (because of its form and lack of interpretation) it should still be considered a major contribution to the teacher education profession at large. It not only is an example of the limitations in its approach to measuring the impact of training, it also has some reliable and provocative data about the performance of student teachers during

training and during the first year of employment. Here, the lack of interpretation is fortunate for the results are so cantankerous, so conflicting with current beliefs about teacher trainee performance, so challenging to current practices and current innovations, that no single interpretation would be believed. Regardless of its many faults as an evaluation study, the Marsh study has generated more evidence that conflicts with current norms than any previous research study into student teacher performance characteristics. Teacher educators should be challenged to make their own interpretations of many of the findings from the Marsh study. The profession can ill afford to disregard the data collected or forget the challenge that the data make to current professional practices and beliefs.

Discussion

As stated previously, the purpose of this paper was to determine why program evaluations are so often ineffective either for helping determine program policy or for generating knowledge for the profession at large. These six evaluation studies of Teacher Corps demonstrate the variety of approaches evaluators have taken in evaluating a dynamic program and their unfortunate ineffectiveness in providing information for Teacher Corps policy or in generating knowledge for the profession. The rather lengthy review of these studies was needed for two reasons. First, they are relatively unknown, but, second, (and most salient to this discussion), it is important to realize that these studies were energetic, highly professional efforts at evaluation. If they have failed as evaluation studies (and they have failed), they have not

done so because of a lack of professional expertise, a lack of attention or effort, a lack of creativity or insight or a lack of respect for the complexity of Teacher Corps and the extent of the evaluation challenge. It is hoped that has been made abundantly clear from the not-so-brief review of these studies.

The following discussion is meant to dwell further upon the irony raised about these studies, that is, that in general they seem to be creative, insightful professional efforts that have generated little useful information.

A question that this discussion will try to answer is: Are the inadequacies of these program evaluations due to not following standard evaluation procedures or are current standard evaluation procedures themselves inadequate for many evaluation purposes? In order to help answer this question, two perspectives of evaluation will be introduced. One perspective is that offered by Alexander W. Astin and Robert J. Panos in their chapter found in the second edition of Educational Measurement and titled "The Evaluation of Educational Programs."[37] The Astin and Panos perspective of evaluation will be used to define what is referred to here as "standard evaluation procedures." Their perspective is used rather than others not only because it can be found in a recent handbook of measurement but also because it is consistent with the views expressed by other major figures in educational program evaluation who are considered to be setting standards for the evaluation profession (e.g. Stufflebeam, Popham, Scriven, Provus). In addition, the Astin and Panos perspective is used because it has one of the widest interpretations of the input, process, output

model that so characterizes standard evaluation perspectives. The Astin and Panos perspective is used because it is a clear, concise and as far ranging a perspective on the input/output model of evaluation as one could expect. It is not a narrow view.

At the same time, although it was meant to synthesize the major characteristics of program evaluation, it is also not the only view. There are views of evaluation that do not accept the input, process, output rationale (e.g. MacDonald, Stake, Elliott, Eisner, Hernandez-Nieto, Glass). There are some perspectives of program evaluation that have been offered as alternatives to standard concepts of evaluation procedures. The alternative perspective used in this analysis comes from Malcolm Parlett's and David Hamilton's paper titled "Illuminative Evaluation."[38] Parlett and Hamilton's view is used not because it is the only alternative available,[39] but because it is a clear, concise, and well written alternative. It is also accessible.

Figure 1 introduces the two evaluation perspectives by describing in phrase form their general views about a) the salient features of educational programs, b) the purposes of program evaluation and c) the nature of the questions that each suggests are appropriate and most crucial to answer in an evaluation study. Even a cursory look at Figure 1 suggests the contrast between these two perspectives. The standard view of educational programs, for example, is that they are transferable, located as an instructional method, a text or a course of study and are designed to produce changes in behavior. In the view of Parlett and Hamilton, on the other hand, any educational program is unique, teaching and learning is defined by the participants, and intentions are often reft with unintended consequences of unintended processes.

Figure 1. Comparison Between Two Evaluation Perspectives

Astin & Panos	Parlett & Hamilton
View of Education Programs	
• a program is an instructional method, single lesson, programmed text, complete course, remedial program, entire school system	• a program is the individuals and group who in turn produce a unique pattern of circumstances, customs, opinions and work styles which suffuse the teaching and learning that are there
• is designed to produce changes in behavior to exposed individuals	• has pervasive operating assumptions
• has some recognized objectives	• is governed by unique teacher characteristics
• is predicated upon belief of the existence of causal relationships between educational objectives and institutional means	• where causal relationships include unintended as well as intended outcomes
• is comprised of outputs, inputs and operations (3 conceptually distinct entities)	• "inputs" include phenomena of crucial educational significance and are difficult to ascertain before instruction (e.g. boredom, interest, concentration, floundering and intellectual dependency)
• a program's objectives are its outputs (e.g. its goals, criteria, outcomes, achievements, dependent variables)	
Purpose of Program Evaluation	
• to provide relevant information about causal relationships between inputs, outputs and operations of the program	• to sharpen discussion, disentangle complexities, isolate the significant from the trivial and to raise the level of debate about the program
• to select those educational operations that are most likely to maximize the student's performance on desired outcomes	• to understand the process occurring within the learning milieu rather than the outcomes
• to determine different effects of educational operations on educational outputs	• to provide an informed account of the innovation
• to produce information for educational decision making based upon clearly stated objectives and the achievement of behavioral manifestations of the objectives	• is likely to increase rather than lessen the sense of uncertainty in education
	• to move beyond helpless indecision or doctrinal assertion in conduct of instructional affairs

Figure 1. Comparison Between Two Evaluation Perspectives (cont'd.)

Astin & Panos	Parlett & Hamilton
Major Questions	
• What was the students' progress or achievement? • What was the impact of the educational program? • Did the chosen alternative produce the expected outcome? • Were the specific intermediate objectives reached? • What were the differences between the students' input and output performances? • How are the outputs best defined and measured? • What are the observed relationships between operations and educational outputs relative to educational inputs?	• How has the program operated? • How has the program been influenced by the various school situations in which it has been applied? • What do the participants (and others directly concerned) regard as its advantages and disadvantages? • What is it like to be participating in a scheme, whether as teacher or student? • What may be the program's most significant features, recurring concomitants and critical processes? • What are the relationships between beliefs and practices?

The purpose of program evaluation, for Astin and Panos, is to provide relevant information about causal relationships between inputs, outputs and operations of the program. The purpose is to identify operations that maximize performance, where performance is the achievement of clearly stated objectives. An alternative purpose, as Parlett and Hamilton suggest it, is to understand the process that occurs rather than lessen the sense of uncertainty one has about cause and effect relationships in the educational process.

These differences in the general outlook of what is important to each of the two views of evaluation are reflected by the kinds of questions each would expect answered in an evaluation study. For Astin and Panos, the questions should include: Did the chosen alternative produce the expected outcome?, and What are the observed relationships between educational operations and educational outputs relative to educational inputs? For Parlett and Hamilton, the questions should include: How has the program operated?, and What are the relationships between beliefs and practices?

When using these two views to analyze the six Teacher Corps evaluation studies, clearly the studies approached their problem as Astin and Panos suggest. All six studies attempted to define the Teacher Corps program in transferable, non-idiosyncratic terms. The RMC I study was so intent upon identifying the objectives of Teacher Corps that that is principally all it did. The Marsh study focused so much upon the causal relationships between inputs, outputs and operations that their most interesting, surprising and trustworthy data on student performance

went mostly unnoticed and uninterpreted. The Abt study was so concerned with its assumed responsibility for determining specific cause-effect relationships between planned operations and outcomes that the reported results from its surveys and quantification of the case studies fairly contradict or at least disregard most of the information contained in those case studies. The Harvard report, concluding that evaluation was logically a nemesis to a political agency like Teacher Corps, never realized the irony that it itself was an evaluation study with a view of education, a set of purposes, questions and procedures that contradicted their own definition of evaluation! In short, the Harvard study approached its own evaluation task in the manner suggested by Parlett and Hamilton (for example, it emphasized pervading operating assumptions and the milieu in which it was formed) but defined evaluation strictly along the lines suggested by Astin and Panos. One can only conclude from this how pervasive was the sense of standard evaluation to that group!

The point to be made here, however, is not only that these evaluation studies were clearly using the perspective of standard evaluation to define their focus, purpose and major questions, but that they were adhering to this perspective very well. Each study, in its own way, showed a creativity and initiative in using these perspectives that can only be admired. The accomplishment of RMC I in specifying intermediate objectives and long range objectives for a young Teacher Corps, the creativity of the mapping sentence designed by the Abt study to relate causal relationships between input, process and output variables, the extent to which both the Marsh and Corwin studies focused upon

student performance as output variables and pursued their a priori questions in terms of program operations, all suggest the extent to which the perspective of standard evaluation was not only used, but also was creatively followed in an attempt to evaluate Teacher Corps.

From the Parlett and Hamilton perspective, however, these studies would not be judged so positively. First, the lack of attention towards understanding the training processes being used by different projects point out that the Parlett and Hamilton perspective was not shared by these studies. Operating assumptions and unintended outcomes were seldom considered by these studies as being important features of the Teacher Corps program. For example, although the Abt study, in its case studies, pointed out fundamental unintended consequences of the Teacher Corps projects they described (e.g. the support of the status quo in institutional racism), the study never highlighted these results nor addressed them through their interpretations or recommendations. The basic questions of Parlett and Hamilton were never asked. How the programs operated or what the relationships were between beliefs and practices are yet to be described. Clearly, these six Teacher Corps studies approached the challenge of evaluating this dynamic and unusual agency from the basic principles of a standard perspective of evaluation rather than from an "illuminative" perspective.

Figure 2 further compares these two perspectives of evaluation. Figure 2 includes the suggestions for a) study design features, b) representative variables, c) data gathering procedures, d) data analysis and e) the nature of the final reports from the Astin and Panos perspective and from the Parlett and Hamilton perspective. The contrast between these two approaches should be apparent.

From the perspective of Astin and Panos, a study design would include early definitions of outcomes, exercise a clear control over salient variables and have a definite statement of the causal relationships between operations and outcomes. From the Parlett and Hamilton perspective, however, an evaluation study would not necessarily be charted in advance, would start with as little control as possible so as to emphasize the day to day reality of the events that are being described and would continually cross-check, re-evaluate and re-design its views of causal relationships. Clearly, the two perspectives differ in what each considers to be an appropriate design for an evaluation study.

Likewise, Figure 2 indicates that the two perspectives differ on a) potential variables to be considered (e.g. the chosen curriculum vs. the program designers' tacit assumptions), b) appropriate data gathering procedures (e.g. direct measurement of concrete outcomes vs. presence at a wide variety of related events) and c) corresponding data analyses (e.g. statistical analyses of relationships vs. observing, inquiring and seeking to explain). As Figure 2 also shows, even the nature of the final reports would differ. A final report from the standard perspective would be designed for program administrators and include information on inputs, outputs and causal interpretations. From an illuminative perspective, on the other hand, a final report would try to present a program as a "recognizable reality" to outsiders with the added criterion that enough evidence be presented (including the critical processes and theoretical principles used) for others to make alternative interpretations.

Figure 2. Comparison of Design and Procedures of Two Evaluation Perspectives

Astin & Panos	Parlett & Hamilton
Study Design Features	
• early definition of outcomes • exercise control over differential student inputs • adequately control variables • replicated simultaneously in several different social settings • statement of causal relationship between educational operations and educational outcomes • existence of two or more independent sets of educational operations for meaningful comparison • comparative data based on a comprehensive taxonomy of educational environments	• course of study not charted in advance • focus on day-to-day reality of settings • progressively clarify and re-define problem areas • continual cross-checking of tentative findings • systematically reduce the breadth of inquiry to give more concentrated attention to the emerging issues
Representative Variables	
• the chosen curriculum • the chosen teaching method • design and location of classrooms • the establishment of certain rules and regulations • students' achievements, knowledge, skills, aptitude for future learning, values, personality, interpersonal relations • treatments, learning strategies, teacher style, instructional techniques • organization of learning experiences manipulative structuring of learning environment	• program designers' tacit assumptions • status differences between participants patterns of attendance, seating, utilization of time and facilities students' boredom, interest, concentration, floundering, intellectual dependency • what work has been done, what participants think of their work (the use and value) • participants' comparisons with past experiences • career ambitions, anxieties • intellectual habits • institutional constraints

Figure 2. Comparison of Design and Procedures of Two Evaluation Perspectives (cont'd.)

Astin & Panos	Parlett & Hamilton
Data Gathering Procedures	
• incorporate a relatively broad and diverse set of potential objectives • direct measurement and observation of concrete outcomes • multiple criteria used to measure achievement of educational objectives • measurement of student inputs and student outcomes (measurement of educational environments much less well developed or certain) • careful planning and judicious selection of variables • objective taxonomic information on several programs	• no data gathering method used exclusively or in isolation • emphasis on observing and interviewing documentation of day-to-day activities • be present at a wide variety of related events • assemble an information profile from: observation, interviews, questionnaires, tests, documentary and background sources • use open-ended comments • compile work diaries, examples of students' assignments • perspective of how the program was regarded by different people before it began
Data Analyses	
• statistical analyses chosen to produce empirical information on the relationships between outputs, inputs and operations • statistical analyses chosen whose assumptions can be satisfied (e.g. simple covariance often not appropriate)	• three stages of analysis: 1) observe, 2) inquire further and 3) seek to explain • continuous weighing of alternative interpretations in light of new information obtained • not relating different test scores, but in accounting for them using the study's findings • cross-check most important findings • use of outside consultants to challenge preliminary interpretations, • play devil's advocate, check open-ended material and develop own interpretations

Figure 2. Comparison of Design and Procedures of Two Evaluation Perspectives (cont'd.)

Astin & Panos	Parlett & Hamilton
Nature of Final Report	
• designed for program administrators and instructors • designed for formulation and refinement of educational theory • include information of the student outputs • include information on the student inputs • all plausible rival interpretations considered before causal interpretations given • replace existing folklore with research information	• present program so that it can be seen as a recognizable reality by others outside the innovation • may increase rather than lessen the sense of uncertainty • document critical research processes • theoretical principles and methodological ground rules discussed and made explicit • criteria for selecting or rejecting areas of investigation spelled out • present evidence in such a way that others can judge its quality and make alternative interpretations

An analysis of the six Teacher Corps evaluation studies from these two perspectives of evaluation provides additional evidence that the studies not only assumed the basic principles of a standard perspective but also performed credibly from that perspective. The variables chosen in the evaluation studies, for example, were consistent with those suggested by Astin and Panos. These variables included an identification of a chosen curriculum (e.g. CTBE - the Marsh Study), a chosen teaching method (e.g. "individualized instruction" - Corwin), the established rules and regulations (e.g. agency guidelines - Corwin and the RMC studies), students' achievements, skills, personality (Corwin and Marsh studies) and treatments or instructional techniques (RMC II and Marsh). Judging by the principles laid out by Astin and Panos, these studies chose the variables "correctly."

As assessment of the choice of variables from the perspective of Parlett and Hamilton, however, would not be so consistently positive. Although the variables identified by the Corwin report would be rated relatively high using Parlett and Hamilton's perspective (since it did focus upon institutional constraints, status differences and career ambitions and anxieties) the two RMC studies would have to be rated very low since none of the variables suggested by Parlett and Hamilton held much importance. Variables like designers' tacit assumptions, status differences between individuals, patterns of utilizing time, student boredom or student intellectual dependency were of little interest to these studies. Thus, from the Parlett and Hamilton perspective, the variables chosen for these studies would generally be inappropriate and consistently be incomplete.

There are also contrasts in the data gathering procedures suggested by these two perspectives on evaluation. Although both, for example, suggest a wide range of data gathering procedures, Astin and Panos emphasize the objectivity of these observations, the measurement of inputs, outputs and (where possible) procedures. Again, along the lines suggested by Astin and Panos, these Teacher Corps evaluations were uncommonly professional. Most incorporated a broad range of potential objectives, directed their measurement to "concrete" outcomes, measured student input and output, used careful planning and judicious selection of variables and, to the best of their capabilities, provided objective taxonomic information on several programs. The recommendations of Parlett and Hamilton that differed from the standard recommendations for data gathering, however, were not incorporated into these studies. Documenting day-to-day activities, for example, or being present at a wide variety of related events were not part of the procedures used to gather data. Thus, although there are commonalities between the data gathering procedures suggested by Astin and Panos and those suggested by Parlett and Hamilton, one would still expect a very positive judgment using the criteria suggested by Astin and Panos and a more negative judgment using the criteria suggested by Parlett and Hamilton.

The data analyses used in these evaluation studies must be considered as practical compromises with the two principles stated by Astin and Panos. That is, the obvious intentions of these studies were a) to produce empirical (statistical) information on the relationships between outputs, inputs and operations and b) to choose statistical

procedures whose assumptions could be satisfied. In practice the compromise with these principles included using available statistical procedures to produce correlations between the chosen variables and later apologize for the extent to which basic assumptions of these procedures were violated. Multiple linear regression and various orthogonal transformations were used, for example, by the Corwin and Marsn studies, to reduce the number of factors and arrive at more interpretable results about their input, output and process variables. Although Corwin seemed to be quite successful in applying these procedures, he also made clear the extent to which basic assumptions such as the additivity and constancy of treatment effects and the lack of interference between experimental units just could not be satisfied given the realities of Teacher Corps and its projects. In fact, the most relevant information reported in Corwin (Chapter Eleven) was reported as conjecture, interpretations and hypothesis. In short, using the criteria of Astin and Panos the data analytic procedures chosen in these studies would have left much to be desired.

Using the criteria of Parlett and Hamilton, it is difficult to arrive at a simple assessment of the data analyses used. Each study eventually realized that it was encountering more than it had anticipated. From the Harvard report, which came to a conclusion that contradicted its original premise (i.e. that standard evaluation was responsible for effective conduct of a national agency), to the RMC, Abt, and Corwin studies, which included stages of further inquiry, some weighing of alternative interpretations and trying to account for the "findings" with other information such as case study information

or informal contacts, the recognition of possible pluralistic interpretations was implicit. From the Parlett and Hamilton perspective, however, it was unfortunate that these interpretive difficulties were not anticipated nor planned for in the study designs. The lack of alternative interpretations or the minimal efforts given to continually alter interpretations as the studies progressed would suggest that the basic principles of analysis offered by Parlett and Hamilton were not followed. Nevertheless, in reading the conclusions of these studies, one would be led to say that these principles of analysis were more tacitly followed as the limitations of statistical results became more apparent. All studies, for example, reported a learned appreciation for the complexities, ambiguities and contradictions represented within those Teacher Corps contexts that were being evaluated.

The nature of the final reports of these six studies consistently followed the principles of Astin and Panos. They included information on student inputs and outputs and clearly were directed at replacing "folklore" with "research information." The attempts to reduce conclusions to specific variables reflected an intent to write to program administrators. If one were to seriously assess the extent to which these reports would have been clearly understood by program administrators, however, these reports would probably receive very low grades. Likewise, as has been suggested previously, except for the Corwin study which used a perspective of organizational theory, there was little refinement or formulation of educational theory.

From the viewpoint of Parlett and Hamilton, the final reports of these six evaluation studies were utter failures. In no way was there

an attempt to present the program as a recognizable reality to others. Evidence was not presented for alternative interpretations. The point of these reports was definitely to lessen (rather than increase) the sense of uncertainty. And, except for the Corwin and Abt fascinations with organizational theory, there were few theoretical principles and methodological ground rules made explicit. In short, although the final reports may have been found wanting in some aspects from the Astin and Panos perspective, they would have been totally unacceptable from the view of Parlett and Hamilton.

Ironically, the study design offered unique problems. Each study, as Astin and Panos suggest, did set out to accomplish an "early" definition of outcomes, devoted obvious energies to stating anticipated causal relationships between these variables and tried to make comparisons upon comprehensive taxonomies of "environments." Each also, however, reported the extent to which they simply could not control the variables they had identified such as student input, training procedures, project setting or project operating characteristics. This lack of control by the evaluators of a complex program such as Teacher Corps is not surprising. Nevertheless, it does point out the difficulties the six studies had in applying certain standard study design features to the evaluation of Teacher Corps projects.

It is also clear, however, that these studies did not use a Parlett and Hamilton perspective in designing their studies. The focus was definitely not on the day-to-day reality of the settings, problem areas were not dwelt upon for clarification and redefinition and continual cross-checking was only in special case studies rather than in

the "final" analyses, interpretations and recommendations. From a Parlett and Hamilton perspective, the design of these studies was not aimed in the right direction. From an Astin and Panos perspective, the aim of the design was there but they fell short of the mark, mostly because of conditions that could not be controlled.

Conclusion

These six evaluation studies performed credibly from the perspective of standard evaluation. The questions that were asked, the variables that were chosen, the data gathering procedures used and the nature of the final reports not only followed many of the criteria suggested by Astin and Panos but did so creatively. Even some of the failures from the Astin and Panos perspective, that is, the lack of control over the variables or the inability to satisfy the assumptions necessary for the statistical analyses chosen suggest the extent to which the basic principles of the standard perspective of evaluation were followed irregardless of the realistic problems encountered. When compared to the Parlett and Hamilton perspective of program evaluation, there can be no doubt that these six large scale evaluations of Teacher Corps clearly applied the standard perspective of evaluation as best they could, given the complexities of the Teacher Corps concepts, goals, guidelines and activities.

In addition to the conclusion that these studies performed credibly according to principles of standard evaluation, there is unmistakable evidence that these studies were not (and could not be) useful to either Teacher Corps, high level agency decision makers or to professional understanding. Major testimony from these studies is not only to

perseverance, hard work and (at times) creativity with which these principles were applied, but also to the general lack of the utility of these principles for generating knowledge about either the program, its effects or lessons about its successes and failures.

From the evidence of these six evaluation studies of Teacher Corps, one is forced to conclude that the failure of these studies is a failure of the methodology and perspective rather than a failure in not following suggested procedures. The principles stated by Astin and Panos are inadequate for their own purposes when evaluating federal agencies such as Teacher Corps.

Where does this leave us? Certainly there is no proof here that an alternative concept of evaluation will work. Nevertheless, the discussion using the Parlett and Hamilton perspective shows that there are perspectives of evaluation that are (as they argue) alternatives to standard perspectives. The conclusion from this review of evaluation studies strongly implies that is is necessary to try out federal program evaluation from perspectives other than those represented by Astin and Panos. Parlett and Hamilton's "illuminative evaluation" is one candidate among many. Other possibilities include MacDonald's notion of "democratic evaluation," Stake's concept of "responsive evaluation," Hernandez-Nieto's suggestion of "process-evaluation" or Bronfenbrenner's concept of "ecological experiments."[40] Considering the amount of professional effort and creativity expended in previous attempts to evaluate Teacher Corps and the corresponding absence of information about the workings of the agency, its local projects, the training procedures

intended, altered and used or the effectiveness of this training, there is no choice but to encourage the use of alternative methodologies in evaluating complex federal programs.

Recommendations

Recommendations from the above conclusion are so obvious that they need hardly be mentioned. These include: a) the need for more flexible evaluation guidelines and principles by the Office of Planning, Budget and Evaluation (OPBE), b) the need for more systematic accounts of alternatives to the current standard perspective of educational evaluation and c) the need to provide more training and experience in those procedures necessary to perform evaluations from alternative perspectives.

In addition to the recommendation that different perspectives of evaluation be practiced and supported in evaluating federal programs, it is recommended that the following suggestions be considered. The point being raised by these suggestions is that too much is taken for granted about federal program evaluation. These suggestions are made to encourage further discussion about the possibilities for a) providing more information that can be useful in determining policy, or b) further illuminating professional and public understanding of educational processes that are being described and assessed.

Suggestion: *Have professionals evaluate the agency who are themselves responsible for determining or carrying out policy on similar programs.*

For example, if Teacher Corps is to be evaluated again, why not include a) teacher representatives, b) school of education personnel, c) state department of public instruction personnel or, maybe d) an international team from similar backgrounds and concerns as part of

the evaluation effort? Part of the rationale for this suggestion is that the most that can be done now is to offer information that allows policy makers to vicariously experience the consequence and meanings of certain policies and activities in order that better (more informed) policy can be made in the future. There are, of course, limits to what evaluators can communicate to "decision makers." It should also be realized, however, that evaluating an agency as complex as Teacher Corps is itself a powerful intervention on the evaluators. The impact of the evaluation process upon the evaluators can be tapped for professional purposes rather than dissipated to private evaluation firms.

Suggestion: *As much effort should be expended in making the reports clear, interesting and useful as there is in collecting, analyzing or interpreting the results.*

Why are so many evaluation reports so difficult to understand and boring as well as useless? The suggestion here is that one reason is the lack of time and effort given to communicating or demonstrating the "findings." Not only is it necessary to identify the audiences and write to them, it is also necessary to choose the most direct way to make the messages understood. One may expect, for example, different kinds of reports for different audiences. The possibilities for the nature of reports may be endless but these possibilities are determined by two factors. One is the imagination of the evaluators (along with their knowledge of the audiences), another is the amount of time, effort and expertise that is made available. Successful communication of results must be the major criterion of evaluation.

Suggestion : Continually (or at least periodically) include client policy makers in on the interpretive procedures.

For an agency like Teacher Corps there are at least three sets of policy makers. One is the legislature which determines the funding levels and mandates the purposes, two is the parent agency (Office of Education) to whom the agency reports, three is the agency personnel themselves, including the director, staff, local (and future) project directors. Each policy level has its own set of questions that need be answered for certain decisions. Often the nature of the questions and decisions can change dramatically over time. The suggestion here is not only to make sure that changes in policy priority and changes in evidence and interpretation be monitored by both evaluators and their clients. It is necessary that each have a realistic perspective of the concerns of the other as well as the limits of the findings.

Suggestion : Conduct federal agency evaluations as interpretive efforts rather than research results.

One of the major conclusions of this review was that most of what is given as "conclusions" is little related to "fact." The results, the instruments, the procedures and data analytic methods to arrive at the results although occasionally interesting (and perhaps "the best there is") are little related to the most useful, helpful and insightful statements. Remember, these are evaluation studies, more precisely, they are efforts to demonstrate the success or failure of certain intervention strategies. Conclusions from these studies do not come directly from the data and data analyses. Thus, the expertise needed to interpret the data is a far different sort than that needed to form conclusions in

tightly controlled educational research. What is suggested here is an approach to the study in the following vein: a) this is what we wanted to know and why we wanted to know it, b) this is how we tried to understand what occurred, and c) these are our interpretations and recommendations from what we have learned. In short, there is no such thing as unambiguous data in these reports. Wisdom, past experience, understanding, creativity and the professional values held by the participants in the evaluation will be the governing factor to the usefulness, "validity" and excitement of the recommendations. Technical expertise of a researcher is simply not enough. Thus, it becomes very important to choose a team of seers and idea people (preferably from different perspectives) to design, gather, interpret, report and recommend.

Suggestion: Consider the needs of the profession for unique information or unusual perspectives and interpretations of their practices along with the needs of policy makers.

Evaluation studies often provide unique opportunities to the profession at large. They offer the profession opportunities to observe natural field settings, especially to observe unusual personal, group or institutional efforts being expended in these settings. The potential for field information that can both shape and re-make professional perspectives is too great not to be taken into account when considering the design of an agency evaluation. Evaluation offers challenges to the professional perspectives normally used to describe natural settings and also challenges how lessons are or can be learned from these settings. Another reality of agency evaluation is that <u>every</u> evaluation report is

an historical document (as well as a policy statement) as soon as it is written. Keeping this in mind, it makes sense to record events, processes, intentions and outcomes (intended and unintended) as historical data describing a particular case rather than as a timeless proof of a universal principle. No matter how information is to be collected, an evaluation study can be designed so that experiences and lessons from the agency can be shared with the profession as well as with those directly responsible for agency activities.

Suggestion : *Support evaluation methods that can capture growth, dynamic interplay, liveliness and impact (as well as the absence of these qualities).*

The ways to capture and interpret dynamic settings must be addressed creatively with few professional restrictions upon what is allowed as "evidence." Currently what is missing from evaluation studies is any sense of the liveliness, growth, periodic failures and subsequent redirections that make up most of the existence of an interventive project. Without ways to capture these features of intervention strategies, there is no reason for participants, interested bystanders, or decision makers to accept the description as real or the interpretation as valid. For example, one common characteristic shared by dynamic features is time, yet time is a non-variable in most evaluation studies (even for those in which "process" is described).

Current events, past failures and immediate needs for more understanding of the information that has been collected leave us with no choice but to search for an alternative evaluation paradigm. As this discussion has emphasized, that alternative must include a greater

understanding of the "black box" of educational processes, must approach "treatment" itself as a problematic, dynamic variable, needs more than a priori definitions and categories of intentions (for both process and outcome) that allows for alternative assumptions about what is to be termed "success" and "failure," and focuses not only upon what are common sense variables but also the extent to which these variables can and cannot be manipulated in specific settings. In short, an alternative is needed that is appropriate for describing, illuminating and interpreting the dynamics of a particular program.

This review of past studies of Teacher Corps has suggested that "better performance" of current standard evaluation procedures cannot be expected to produce better results. It must also be realized, however, that there is no conclusive evidence that an alternative set of procedures will be more useful either to policy determination or to professional understanding. The Parlett-Hamilton perspective has only suggested that there exist perspectives and accompanying methodologies that offer different criteria for excellence and different possibilities for success. We will never know more than that alternatives exist until those alternatives, too, can be subjected to the difficult challenges of federal agency evaluation. Considering the current need for more appropriate methodologies, the possibilities for learning much more about what would constitute successful programs and the existence of alternative evaluation designs, data collection and data analysis procedures, our failure to employ these alternative designs on at least an experimental basis is gross professional negligence. In conclusion, this critical analysis of past Teacher Corps evaluations suggests that we have used an extremely arid approach to measure our Springs with.

Footnotes

1. This is a stanza from a poem, no. 23, "voices to voices, lips to lips", in e.e. cummings (1923).

2. For example, see Apple (1974), Karier (1974), Novack and Gilbert (1975), Ross and Cronbach (1976), Stake (1967) and Young (1971), who question the ideological or theoretical efficacy of standard evaluation procedures.

3. For example, Jackson (1975) critically analyzes Coleman data on school integration.

4. Teacher Corps: Past or Prologue?, a report by the National Advisory Council on Education Professions Development, George Arnstein, executive director (1975) offers a concise introduction to Teacher Corps.

5. The Marsh and Lyons study (1974) is one example.

6. For example, the Corwin study, Reform and Organizational Survival (1973) was funded jointly by NEA and the Ford Foundation.

7. For example, the Cort and O'Keefe study (1968) was funded by Teacher Corps.

8. Agencies such as Abt Associates, Incorporated, Resource Management Corporation and Pacific Training and Technical Assistance Corporation have taken part in evaluations of Teacher Corps.

9. See Teacher Corps: Past or Prologue? (op cit).

10. As evidenced in the Federal Register, December 17, 1975, Teacher Corps projects became "demonstration" projects.

11. Davis and Amershek (1969), Haberman (1971) and Dusalt (1970) are teacher educators who were identifying the need for a greater knowledge base and more empirical research into teacher education at the same time that these evaluation studies were being conducted.

12. Astin and Panos (1971).

13. Parlett and Hamilton (mimeo in 1974, published in 1976).

14. For a smooth and continuous reading, footnotes will only be made in these descriptions of evaluation studies for direct quotes.

15. In 1975, a technical assistance project was funded for providing Teacher Corps with a Management Information System (MIS).

16. Chapter IV: The Role of Evaluation in the Bureau of Educational Development, p. 109.

17. ibid. p. 155.

18. ibid. p. 156.

19. Full-Scale Implementation of a Process Evaluation System for Programs of NCIES, p. 40.

20. ibid. p. 40.

21. Innovation and Change, Volume I, p. 1.

22. ibid. Volume II, p. 32.

23. Corwin (1973), p. 11.

24. ibid. p. 116.

25. ibid. pp. 116-117.

26. Federal Role In School Reforms From Sociological and Educational Perspectives. Edited by G.T. Fox (1974).

27. Since the Marsh study, the closest effort to generating information about teacher training practices of Teacher Corps projects was the 1975 CMTI Impact Study (Fox, Grant, Popkewitz, Romberg, Tabachnick and Wehlage).

28. Marsh and Lyons, Follow-Up Study, p. 15.

29. ibid. p. 16.

30. ibid., for example, pp. 16, 17, 18.

31. ibid. p. 18.

32. See Fox, mimeo (January, 1974).

33. It reflected, for example, both the views of Scriven (1967) and Stufflebeam et al (1971) that the major responsibility of program evaluation is to attempt to determine causal relationships between variables.

34. Marsh and Lyons, Volume I, p. 166.

35. ibid. p. 1.

36. Marsh and Lyons, Follow-Up Study, p. 46.

37. See Astin and Panos (1971).

38. See Parlett and Hamilton (1976).

39. See Carini (1975), Eisner (1975), MacDonald (1974), Stake (1972).

40. These views are explicated in MacDonald (1974), Stake (1972), Hernandez-Nieto (1976), Bronfenbrenner (1976).

REFERENCES

1. An Assessment of Teacher Corps. 2 volumes. Bethesda, Md.: RMC Incorporated (Robert Crosby, Director), August, 1970. Grant no. OEC-0-099018-4647.

2. Apple, Michael. "The Process and Ideology of Valuing in Educational Settings." In Apple, Subkoviak and Lufler. Educational Evaluation: Analysis and Responsibility. Berkeley: McCutchen Publishing Corporation, 1974.

3. Astin, Alexander and Panos, Robert. "The Evaluation of Educational Programs." In Thorndike, Robert L. Educational Measurement, Second Edition. Washington, D.C.: Averiean Council on Education, 1971. pp. 733-751.

4. Bloom, B., Hastings, J. and Modans, G. Handbook on Formative and Summative Evaluation of Student Learning. New York: McGraw-Hill, 1971.

5. Bronfenbrenner, Uri. "The Experimental Ecology of Education." Educational Researcher 5 (9) October 1976, pp. 5-15.

6. Brugelman, Hans. "Towards Checks and Balances in Educational Evaluation: On the Use of Social Control in Research Design." In MacDonald, B. and Walker, R. Innovation, Evaluation, Research and the Problems of Control. East Anglia, England: SAFARI PROJECT Interim Papers. Centre for Applied Research in Education, 1974.

7. Bussis, A., Chittenden, E.A. and Amarel, M. Methodology in Educational Evaluation and Research. Unpublished manuscript. Princeton, NJ: Educational Testing Service, 1973.

8. Campbell, Donald T. "Degrees of Freedom and the Case Study." Comparative Political Studies. 8 (2), July 1975, pp. 178-193.

9. Campbell, Donald. "Reforms As Experiments." In Struening, E. and Guttentag, M. Handbook of Evaluation Research, Volume 1. Beverly Hills: Sage Publications, Inc., 1975.

10. Carini, P.F. Observation and Description: An Alternative Methodology for the Investigation of Human Phenomena. Grand Forks, ND: University of North Dakota, North Dakota Study Group on Evaluation, Center for Teaching and Learning, 1975.

11. Chapter IV: The Role of Evaluation in the Bureau of Educational Development. "The Teacher Corps: A Case Study in Evaluation." Boston: Center for Educational Policy Research, Harvard University, December, 1971.

12. Cicirelli, V., Evans, J., Schiller, J. "The Impact of Head Start, A Reply to the Report Analysis." Harv. Educ. Res. 40, 1970, 105-129.

13. Coleman, James. "Recent Trends in School Integration." Educational Researcher, 4 (7), 1975, pp. 3-12.

14. Coleman, James, et al. Equality of Educational Opportunity. Washington, D.C.: U.S. Government Printing Office, 1966.

15. Cort, H. Russel, Jr., and O'Keefe, Ruth Ann. "Teacher Corps - Two Years of Progress and Plans for the Future." Report prepared under a contract from the Office of Education, Washington, D.C., October, 15, 1968. Monograph.

16. Corwin, Ronald G. Reform and Organizational Survival: The Teacher Corps as an Instrument of Educational Change. New York: Wiley-Interscience, 1973.

17. Corwin, Ronald. What Difference Has the Teacher Corps Made? Longitudinal Studies of the Teacher Corps Program. Report submitted to the Ford Foundation through the NEA. September 1973.

18. cummings, e.e. 100 Selected Poems. New York: Grove Press, 1923.

19. Davis, Don and Amershek, Kathleen. "Student Teaching." In Ebel, Robert L. Encyclopedia of Educational Research, 4th edition. New York: MacMillan and Co., 1969.

20. Dusalt, Gilles. A Theory of Supervision in Teacher Education. New York: Teacher College Press, 1970.

21. Eisner, E. "Emerging Models for Educational Evaluation." School Review. 80, 1971, pp. 573-590.

22. Eisner, E. "The Perceptive Eye: Toward the Reformation of Educational Evaluation." Stanford Evaluation Consortium, Department of Education, Stanford University. Occasional Paper, December, 1975.

23. Federal Register. 40 (243) Wednesday, December 17, 1975. pp. 17-43.

24. Fox, G.T. "Personal Notes and Reactions to Marsh Report" and "Revision of Variables of Marsh Report." Mimeo, University of Wisconsin-Madison, January, 1974.

25. Fox, G.T. (ed.) Federal Role in School Reform From Sociological And Educational Perspectives. Madison, Wisconsin: University of Wisconsin, 1974. Published through a grant from Teacher Corps Projects, Office of Education.

26. Fox, G.T., Grant, C., Popkewitz, T., Romberg, T., Tabachnick, B. and Wehlage, G. 1975 CMTI Impact Study Technical Reports Numbers 1-21. Madison, Wisconsin: University of Wisconsin-Madison. September 1976. Prepared as part of Teacher Corps, Office of Education contract no. 300-75-0100 (University of Nebraska-Omaha).

27. Full-Scale Implementation of a Process Evaluation System for Programs of NCIES. Report UR-172, Volume 1, and Part G. Volume 2. Bethesda, Md.: RMC Incorporated (Robert Crosby, Director), November, 1972.

28. Glass, Gene. "A Paradox About Excellence of Schools and the People in Them" Educational Researcher. March 1975, pp. 9-13.

29. Glass, Gene (ed.) Evaluation Studies Review Annual. Volume I. Beverly Hills: Sage Publications, 1976.

30. Gottman, J.M. and Clasen, R.E. Evaluation in Education. A Practitioner's Guide. Itaska, Ill.: F.E. Peacock Publishers, Inc., 1972.

31. Guba, E. and Stufflebeam, D. Evaluation. The Process of Stimulating, Aiding and Abetting Insightful Action. Monograph Series In Reading Education No. 1. Bloomington: Indiana University, 1970.

32. Haberman, Martin. "Twenty-three Reasons Universities Cannot Educate Teachers." The Journal of Teacher Education. V XXII, No. 2, Summer, 1971.

33. Hernandez-Nieto, Rafael. Process-Development: An Alternative Approach for Curriculum Research in Mathematics. Madison, Wisconsin: Wisconsin Research and Development Center for Cognitive Learning. Working Paper No. 197. November, 1976.

34. House, Ernest. "Justice in Evaluation." In Glass, G. Evaluation Studies Review Annual, Volume I. Beverly Hills: Sage Publications, 1976.

35. Innovation and Change: A Study of Strategies in Selected Projects Supported by NCIES. 4 volumes. Cambridge, Mass.: Abt Associates, December, 1972. Grant no. OEC-9-71-3714.

36. Jackson, Gregg. "Reanalysis of Coleman's Recent Trends in School Integration." Educational Researcher. 4(10) 1975, pp. 21-25.

37. Karier, Clarence. "Ideology and Evaluation: In Quest of Meritocracy." In Apple, Subkoviak and Lufler. Educational Evaluation: Analysis and Responsibility. Berkeley: McCutchen Publishing Corporation, 1974.

39. MacDonald, Barry. "Evaluation and the Control of Education." In MacDonald, B. and Walker, R. Innovation, Evaluation, Research and the Problem of Control. East Anglia, England: SAFARI PROJECT Interim Papers, Centre for Applied Research in Education, 1974.

40. MacDonald, Barry and Parlett, Malcolm. "Re-thinking Evaluation: Notes from the Cambridge Conference." Cambridge Journal of Education, 3 (2), Easter Term, 1973. pp. 74-82.

40. Marsh, David and Lyons, Margaret. A Study of Teacher Training at Sixth-Cycle Teacher Corps Projects. Volume I. Methodology and Findings. July, 1974; A Study of the Effectiveness of Sixth-Cycle Teacher Corps Graduates. Phase II. October, 1974; Teacher Corps Graduates in Highly Innovative Roles in Education: A Follow-Up Study. Phase II. February, 1975. Pacific T and TA Corporation. U.S. Department of Health, Education and Welfare, Office of Education, Office of Planning, Budgeting and Evaluation. Contract no. OEC-0-73-5174.

41. Novack, Mark and Gilbert, Sid. "Paradigms and Policy." Paper delivered at 1975 AERA Convention, April, 1975.

42. Parlett, M. and Hamilton, D. "Evaluation as Illumination: A New Approach to the Study of Innovatory Programs." In Glass, Gene. Evaluation Studies Review Annual, Volume I. Beverly Hills: Sage Publications, 1976. pp. 140-157.

43. Popkewitz, Thomas. "Legitimatizing Failure: Teacher Corps as a Reform." School Review 84 (1).

44. Provus, M. Discrepancy Evaluation. Berkeley: McCutchen Publishing Corporation, 1971.

45. Ross, Lee and Cronbach, Lee (ed.) "Handbook of Evaluation Research, Essay Review." Educational Researcher 5 (10) November, 1976, pp. 9-19.

46. Scriven, Michael. "The Methodology of Evaluation." In Perspectives of Curriculum Evaluation. AERA Monograph Series on Curriculum Evaluation. No. 1. Chicago: Rand McNally, 1967.

47. Scriven, Michael. "Pros and Cons About Goal Free Evaluation." Evaluation Comment 3 (4) 1972, pp. 1-14.

48. Scriven, Michael. "Maximizing the Power of Causal Investigations: The Modus Operandi Method." In Glass, G. Evaluation Studies Review Annual, Volume I. Beverly Hills: Sage Publications, 1976.

49. Stake, Robert. "The Countenance of Educational Evaluation." Teachers College Record. 68, 1967, pp. 523-540.

50. Stake, Robert. "An Approach to the Evaluation of Instructional Programs (Portrayal vs. Analysis)." Paper presented at 1972 AERA Convention. Chicago, April, 1972.

51. Stufflebeam, D., Foley, W., Gephart, W., Guba, E., Hammond, R., Merriman, H. and Provus, M. Educational Evaluation and Decision Making. Bloomington, Ind.: Phi Delta Kappa National Study Committee on Education. 1971.

52. Suchman, E. Evaluation Research: Principles and Practice in Public Service and Social Action Programs. New York: Russell Sage Foundation 1967.

53. Teacher Corps: Past or Prologue? A Report with Recommendations, by the National Advisory Council on Education Professions Development. Walter Tice, Chairman; George Arnstein, Executive Director. Washington, D.C. July, 1975.

54. Weiss, Carol. "The Politicization of Evaluation Research." Journal of Social Issues. 6 (4), 1970, pp. 57-68.

55. Young, Michael (ed.) Knowledge and Control. London: Collier-MacMillan, 1971.